Office of Technology Assessment

Congressional Board of the 98th Congress

MORRIS K. UDALL, *Arizona, Chairman*

TED STEVENS, *Alaska, Vice Chairman*

Senate

ORRIN G. HATCH
Utah

CHARLES McC. MATHIAS, JR.
Maryland

EDWARD M. KENNEDY
Massachusetts

ERNEST F. HOLLINGS
South Carolina

CLAIBORNE PELL
Rhode Island

House

GEORGE E. BROWN, JR.
California

JOHN D. DINGELL
Michigan

LARRY WINN, JR.
Kansas

CLARENCE E. MILLER
Ohio

COOPER EVANS
Iowa

JOHN H. GIBBONS
(Nonvoting)

Advisory Council

CHARLES N. KIMBALL, *Chairman*
Midwest Research Institute

EARL BEISTLINE
University of Alaska

CHARLES A. BOWSHER
General Accounting Office

CLAIRE T. DEDRICK
California Land Commission

JAMES C. FLETCHER
University of Pittsburgh

S. DAVID FREEMAN
Tennessee Valley Authority

GILBERT GUDE
Congressional Research Service

CARL N. HODGES
University of Arizona

RACHEL McCULLOCH
University of Wisconsin

WILLIAM J. PERRY
Hambrecht & Quist

DAVID S. POTTER
General Motors Corp.

LEWIS THOMAS
*Memorial Sloan-Kettering
Cancer Center*

Director

JOHN H. GIBBONS

This is an OTA Background Paper that has neither been reviewed nor approved by the Technology Assessment Board

J952-39

THE IMPACT OF RANDOMIZED CLINICAL TRIALS ON HEALTH POLICY AND MEDICAL PRACTICE

BACKGROUND PAPER

AUGUST 1983

OTA Background Papers are documents containing information that supplements formal OTA assessments or is an outcome of internal exploratory planning and evaluation. The material is usually not of immediate policy interest such as is contained in an OTA Report or Technical Memorandum, nor does it present options for Congress to consider.

CONGRESS OF THE UNITED STATES
Office of Technology Assessment
Washington, D. C. 20510

Recommended Citation:

The Impact of Randomized Clinical Trials on Health Policy and Medical Practice: Background Paper (Washington, D.C.: U.S. Congress, Office of Technology Assessment, OTA-BP-H-22, August 1983).

Library of Congress Catalog Card Number 83-600575

For sale by the Superintendent of Documents,
U.S. Government Printing Office, Washington, D.C. 20402

Preface

Rational decisions at all levels in health care—from Federal Government policymaking to the treatment of a single patient by a physician—require sound information. Randomized clinical trials (RCTs), a family of clinical experimental designs, provide the highest quality of evidence for the efficacy and safety of medical technologies.

The Office of Technology Assessment (OTA) has a longstanding interest in the tools of medical technology assessment and decisionmaking. Previous OTA reports focusing on the information necessary and available for these activities have discussed the role of RCTs in particular. RCTs fill an obvious need for information yet their impact in health care has remained largely undocumented. This background paper was initiated by OTA to bring together the literature and current views about the actual and potential role of RCTs in decisionmaking about medical technologies.

OTA background papers are prepared by OTA staff and drafts are reviewed by interested individuals and organizations. This paper was written by Hellen Gelband. Thomas Chalmers and Henry Sacks prepared an annotated bibliography that provided material for chapter 5. The Health Program Advisory Committee reviewed the draft; those individuals acknowledged in appendix B either provided information during the course of the study, reviewed the draft report, or did both.

Health Program Advisory Committee

Sidney S. Lee, *Committee Chair*
Michael Reese Hospital and Medical Center

Stuart H. Altman*
Florence Heller School
Brandeis University

Carroll L. Estes
Department of Social and Behavioral Sciences
School of Nursing
University of California, San Francisco

Rashi Fein
Department of Social Medicine and Health Policy
Harvard Medical School

Melvin A. Glasser
Health Security Action Council
Committee for National Health Insurance

Patricia King
Georgetown Law Center

Joyce C. Lashof
School of Public Health
University of California, Berkeley

Margaret Mahoney
The Commonwealth Fund

Frederick Mosteller
Department of Health Policy and Management
School of Public Health
Harvard University

Mitchell Rabkin*
Beth Israel Hospital
Boston, Mass.

Dorothy P. Rice
Department of Social and Behavioral Sciences
School of Nursing
University of California, San Francisco

Richard K. Riegelman
School of Medicine
George Washington University

Walter L. Robb
Medical Systems Operations
General Electric
Milwaukee, Wis.

Frederick C. Robbins
Institute of Medicine

Rosemary Stevens
Department of History and Sociology of Science
University of Pennsylvania

---

*Until April 1983.

OTA Staff for The Impact of Randomized Clinical Trials on Health Policy and Medical Practice

H. David Banta, *Assistant Director, OTA*
Health and Life Sciences Division

Clyde J. Behney, *Health Program Manager*

Hellen Gelband, *Project Director*

Virginia Cwalina, *Administrative Assistant*
Ann Covalt, *Editor*
Jennifer Nelson, Secretary
Mary Walls, *Secretary*

Contractors

Thomas C. Chalmers and Henry Sacks
The Mount Sinai Medical Center
Annotated Bibliography on the Impact of RCTs on Medical Practice

OTA Publishing Staff

John C. Holmes, *Publishing Officer*
John Bergling   Kathie S. Boss   Debra M. Datcher   Joe Henson
Glenda Lawing   Linda Leahy   Donna Young

Contents

Chapter *Page*

1. INTRODUCTION 3
    Background 5
    A Brief History of the RCT in Medicine 5
    A Description of the Method 7
    The Uses of RCTs 9
    The Role of the Physician in RCTs 11
    The Diffusion of Medical Technology 12

2. FUNDING FOR CLINICAL TRIALS 19
    Trends in Funding Clinical Trials 20
    The National Institutes of Health 21
    The Alcohol, Drug Abuse, and Mental Health Administration 24
    The Veterans Administration 25
    The Department of Defense 25
    Health Insurers and Support RCTs 26

3. RCTs AND HEALTH POLICY 31
    Drug Regulation 34
    Regulation of Medical Devices 34
    RCTs and Cost-Effectiveness Analysis 34
    RCTs and Medicare Coverage 35
    The Potential Impact of RCTs on Third-Party Payers 37

4. FACTORS AFFECTING THE IMPACT OF RCTs ON MEDICAL PRACTICE 41
    Objections to RCTs 41
    Alternatives to RCTs 44
    Characteristics of RCTs That Affect Their Impact 46

5. THE IMPACTS OF CLINICAL TRIALS ON MEDICAL PRACTICE 61
    RCTs and Concordance With Medical Practice 62
    RCTs in Cancer Research 63
    Cardiovascular Disease 69
    Surgery 78
    RCTs in Other Fields 80

6. IMPROVING THE IMPACT OF RANDOMIZED CLINICAL TRIALS 83
    Improving the Quality of RCTs 83
    Improving the Dissemination and Use of RCT Results 85
    Improving the Assessment of Medical Technologies 87

Page
    The Use of RCTs in Policy Decisions 87
    Improved Use of RCTs in Specific Fields 88
    Conclusion 90

Appendix A. Acronyms and Glossary 93
Appendix B. Acknowledgments 96

References 101

Tables

Table No. *Page*

1. Number and Amount of Support for NIH Supported Clinical Trials Active in Fiscal Year 1979, by Institute for Type of Intervention 19
2. NIH Support for Clinical Trials, Fiscal Year 1979 22
3. Studies Required in FDA's Premarketing Drug Approval Process 23
4. ADAMHA Treatment Assessment Research Fiscal Year 1982 Expenditures 24
5. Office of Health Technology Assessment Report Series 1982 36
6. Distribution by Sites of the 945 Trials Registered at the International Union Against Cancer Information Office, and Related Incidence Rates 67

Figures

Figure No. *Page*

1. The National Heart, Lung, and Blood Institute's Clinical Trial Decision Process 70

Boxes

Page

A. The Society for Clinical Trials 7
B. The National Eye Institute 23
C. Institute for Health Care Evaluation 38
D. Gastric Freezing 47
E. The University Group Diabetes Program 49
F. Hyperbaric Oxygen Treatment for Cognitive Deficits in the Elderly 64
G. The Hypertension Detection and Followup Program 73

1.

Introduction

1.
Introduction

The Office of Technology Assessment (OTA) has had a longstanding interest in the use of randomized clinical trials (RCTs). The OTA report *Assessing the Efficacy and Safety of Medical Technologies* (225) discusses the advantages and disadvantages of RCTs and puts forward a number of policy alternatives for identifying technologies in need of assessment, stimulating clinical trials, and disseminating information derived from them. *The Implications of Cost-Effectiveness Analysis of Medical Technology* (229) discusses the value of RCTs in cost-effectiveness analyses, and notes that information derived from RCTs is not available on many technologies. *Strategies for Medical Technology Assessment* (234) concludes that RCTs are the "definitive experimental method for evaluating the efficacy or health benefits of a technology." Other OTA assessments and case studies in some way use or discuss the results of RCTs (e.g., case studies for *The Implications of Cost-Effectiveness Analysis of Medical Technology*, 1978-1982; *A Review of Selected Federal Vaccine and Immunization Policies*, 1979; *Technology Transfer at the National Institutes of Health*, 1982; *Postmarketing Surveillance of Prescription Drugs*, 1982).

OTA's continuing interest in RCTs led to the question that this study posed: What has been the impact of RCTs on health policy and medical practice? This study is based largely on a review of the literature concerning the history of RCTs and their support, their use in health policymaking, and their influence on medical practice. This review has been supplemented by discussions with policymakers and medical and health specialists with particular interests in RCTs.

The remainder of this chapter contains background material about RCTs and a brief discussion of the diffusion of medical technologies. Chapter 2 covers the funding of RCTs and some nonrandomized clinical trials. The current and possible future uses of RCTs in health policymaking are discussed in chapter 3. Chapter 4 looks at criticisms of and alternatives to RCTs, and the characteristics of RCTs that appear to influence their impact. Chapter 5 reviews the literature specifically about the impacts of RCTs on medical practice. Suggestions for strengthening the impact of RCTs are brought together in the last chapter.

In this paper, "medical technologies" include drugs, devices, and medical and surgical procedures. The organizational and supportive systems through which medical care is provided are part of medical technology in its broadest sense, but they are not discussed here in detail.

Drugs, devices, and procedures are used to diagnose, treat, and prevent disease, and to promote health. Diagnosis usually involves tests and procedures, often using specific medical devices. Treatments may include the use of drugs, devices, and procedures. Disease prevention is traditionally broken down into the categories of primary, secondary, and tertiary prevention. Primary prevention is aimed at avoiding disease altogether. Most vaccines, for instance, are considered primary prevention. Secondary prevention consists of strategies to detect disease in its early stages of development, with the hope of improving patient outcome. Many screening programs, e.g., for breast cancer, are examples of secondary prevention. Tertiary prevention attempts to arrest further deterioration in individuals who suffer later stages of disease. RCTs can be used in evaluations of all types of disease prevention.

RCTs are experiments that test the safety and efficacy of medical technologies. An "experiment" more generally has been defined as "[t]he planned manipulation of material, subjects, or processes by the experimenter, in order to establish a cause-effect relation or a rule (model) for the variation of observations" (151).

In this century, RCTs have replaced anecdotal evidence as the standard for evaluating medical technologies. The development and increasing use of RCTs in evaluating medical interventions is not an isolated phenomenon, but rather part of a broader trend. Experimental methods are increasingly used in studying all types of human problems. In or out of the clinical setting, the randomized trial is the strongest tool available across a spectrum of research topics (56,198). For example, the testing and evaluation of social interventions using randomized designs forms the basis for the growing field of social experimentation. Social and medical issues meet in health services research in evaluating interventions that are not medical technologies, but that are applied in clinical settings. For example, in an innovative program at Cleveland Metropolitan General Hospital researchers have conducted randomized trials on the effect on physicians' ordering of tests when they are provided information or education (168). McGhan and colleagues (148) report a randomized trial comparing pharmacists and technicians as dispensers of prescriptions for ambulatory patients. The use of randomized trials in this field will undoubtedly grow, as it could greatly contribute to the efficient provision of health services. While the study designs in this field are identical or similar to those used to test medical technologies, these studies will not be discussed in detail in this paper.

In clinical settings, RCTs occupy a niche at one end of the spectrum of biomedical research. At the other is found untargeted basic research in biological processes, moving toward preclinical and clinical research and the development of medical technologies for specific diseases. The RCT is a method for testing the efficacy and safety of such technologies. The reason for conducting an RCT should be a sound hypothesis about the technology in question. Fisher (73) notes that the significance of preclinical laboratory research and of clinical trials in fact depend on each other:

> Until a proper clinical test is carried out, no matter how promising a line of investigation seems to be it remains just that, a promise. Clinical research, on the other hand, without a firm biological basis acquired from laboratory investigation is apt to be nothing more than product testing.

Like other kinds of experiments, the RCT compares the effect of an intervention (a medical technology) on one group of people with the fate of a "control" group, which is not subject to the intervention but is otherwise similar to the "experimental" group. RCTs are distinguished from other kinds of comparative studies in that individuals are randomly assigned to these different groups. "Random" does not mean "haphazard" in this case, but rather that individuals are assigned with equal probability to the experimental or the control group.

Randomization is crucial in allowing certain statistical inferences about the experiment's outcome. Random allocation eliminates overt and covert biases in the assignment of patients to treatments. Patients with particular medical characteristics are not determinedly placed more frequently in any one group. Differences in the outcomes of the groups can thus be attributed to the intervention, within the limits of statistical probability.

In other comparative studies, groups are formed by methods other than randomization. But experimenters may be biased in selecting the members of these groups because, consciously or unconsciously, they favor some particular outcome. Such bias would of course compromise the conclusions about why any difference is observed between the groups. Other kinds of epidemiologic and evaluative studies can provide valuable information, though they cannot replace RCTs. See *Strategies for Medical Technology Assessment* (234) for information about the role of other study designs in assessing medical technologies.

The design and execution of RCTs may benefit from prior nonrandomized clinical studies, such as case reports and retrospective analyses of clinic records. "Suggestive evidence" from these sources may provide the justification for carrying out an RCT, and indicate patients most likely to benefit from the technology. The suggestive evidence that "lumpectomy" (removing only a tumor and small amount of tissue) might be effective in treating breast cancer came from retrospective examination of clinic records. An RCT based on that evidence confirmed the value of lumpectomy (188).

Further details about the rationale and methods of RCTs are described in later sections of this chapter.

BACKGROUND

For as long as medical care has been given, people have been concerned about its effects. Does a given treatment cure, prevent, or ameliorate a condition, and what are its other effects, beneficial and detrimental? Nevertheless, specific questions about a treatment's efficacy and safety, not to mention cost effectiveness, have not always been explicit, and attempts to answer them even less so. Concern about the effects of medical interventions has been heightened by three developments in recent decades: the development of more powerful medical technologies; the availability of more effective tools to evaluate them, e.g., the RCT; and the rapidly increasing costs of health care.

During the latter half of the 19th century, quantitative evaluation led to the abandonment of a substantial number of therapies, with no effective therapies to replace them. Major breakthroughs in medical treatment and disease prevention began in the late 19th century and continued through the 1930's and 1940's, brought about by greater understanding of infectious diseases. The advances were obvious, and confidence in medicine ran high. The successes in overcoming many infectious diseases made chronic diseases the major causes of sickness and death in developed countries, and led to new kinds of medical interventions. As success stories became fewer and less dramatic, uncertainty arose again about the value of medical practices.

The rising cost of medical care is one of the most pervasive issues in health care. The development and analysis of strategies to control costs is an area of research itself (see, e.g., 230). New technologies in particular contribute to the rise in both capital costs (e.g., for the new generation of diagnostic imaging equipment such as computed tomography scanners and nuclear magnetic resonance imagers) and health manpower costs (e.g., for intensive care units and complex surgical procedures). Another fact of economic importance is that many technologies can be widely disseminated and used. Imaging, for example, is important in a wide range of medical practice, and new treatments for heart disease address the most frequent chronic disease and cause of death in this country.

The combined concerns for the safety and efficacy of medical practices and for the rising costs of health care together impel the need for rational decisionmaking to avoid what does not work or is unsafe and to get the most for health care dollars. Such decisionmaking depends on information that compares the safety and efficacy of competing technologies. The best method of gathering such information is the RCT.

It has been estimated that between 10 and 20 percent of all current medical procedures have been shown efficacious in controlled trials (225). While it is not possible or desirable to evaluate all medical practices with RCTs, the method could be used much more in evaluating new technologies, in evaluating new applications of existing technologies and in evaluating practices that have long been used but that are still of questionable value (e.g., hysterectomy for some indications).

A BRIEF HISTORY OF THE RCT IN MEDICINE

RCTs are a product of this century, but their forerunners in evaluating "health technologies" reach back at least to Biblical times and in all probability much earlier. An essential element of RCTs, the use of a control group, is related in the Book of Daniel (ch. 1). Daniel was among those children of Israel "in whom was no blemish, but well favored, handsome and skillful in all wisdom" who were chosen to be readied to serve Nebuchadnezzar, the conquering king. Placed in the charge of the prince of the eunuchs, the children were to be fed the king's meat and wine. Daniel, not wanting to be defiled by the diet, asked of the eunuch that he and his three companions from Judah be given pulse (a type of pea) and water instead. The eunuch was afraid he would be blamed for the poor condition of the boys that he thought would certainly result from such nutrition. Daniel convinced him to give them pulse and water for 10 days:

> Then let our countenances be looked upon before thee, and the countenance of the children that eat of the portion of the king's meat.

Ten days later Daniel and his companions were judged "fairer and fatter in flesh" than the other children, and the impact of the trial was immediate and direct. From then on, all the children were nourished on pulse and water.

Careful observation and the use of comparison groups have marked advances for human wellbeing since Daniel's time. Only careful evaluation satisfies healthy scientific skepticism about the value of new technologies. Unfortunately, the need for experimentation is not universally acknowledged, and there are undoubtedly those in medicine today who subscribe to an updated version of the reasoning of a respected 17th century physician: given irrefutable evidence that blood circulates, he replied: "Experiments irritate nature. When nature is irritated it acts otherwise than when it is left alone. Therefore, experiments prove nothing" (94). Nonetheless, progress has been made.

James Lind, in his famous 1747 experiment, compared six treatments for the prevention of scurvy. A full 150 years after the treatment was first suggested in print, he confirmed citrus fruits as a successful prophylaxis (136). The impact of the trial was further delayed: it was 40 years before the British navy required that citrus fruits be carried on ships at sea (40).

The tradition of careful observation and comparison was joined in this century with quantitative methods, to produce modern experimental design (151). In the 1920's and 1930's, R. A. Fisher developed methods for statistical inference based on random allocation, which he applied to his agricultural experiments. Fisher led the way for the medical application of randomization and the statistical methods reliant on random allocation.

The value of knowing which was in fact the first "true RCT" is debatable, but the history is interesting. A. B. Hill was the first major advocate for RCTs in England, where he carried out a trial of patulin against the common cold in 1944 (175) and a trial of streptomycin therapy for tuberculosis, begun in 1946 (161). W.G. Cochran was the earliest strong proponent of RCTs in this country. Some contend that a trial of therapy for tuberculosis published in 1931 by Amberson and colleagues (1) qualifies as the first RCT. In their trial, the control and treatment groups were closely matched on various clinical dimensions, with the choice of which group would get the experimental treatment decided by the flip of a coin. They clearly recognized the value of unbiased allocation, but not the importance of randomization for valid statistical evaluations. Hill, on the other hand, clearly had emphasized randomization (141). Whether Amberson or Hill conducted the "first RCT" is thus a question of whether the experimenter's full awareness of its principles are included in the definition.

The present study concerns the modern RCT, which began with the randomized allocation to treatment groups in clinical settings. This procedure was introduced around the middle of this century at about the same time as the modern generation of drugs, including antibiotics, and vitamins, and other therapeutic measures were developed, demanding standards for evaluation. Adopted initially to evaluate drugs and vaccines, the RCT still enjoys its widest use in that area, its use in evaluating medical procedures and devices developing later and more slowly. The move from using the RCT in evaluating therapies and preventive interventions for acute diseases, to its use in treating and preventing chronic diseases occurred first during the late 1950's in tests of new treatment regimens for leukemia. In the 1960's, RCTs were employed in developing treatment regimens for other chronic diseases, notably cardiovascular diseases. They have also been used in testing diagnostic techniques (e.g., mammography to detect breast cancer), though still infrequently.

The use of RCTs has shown steady growth. In a random sample of articles from general medical journals, no RCTs were reported in 1946, while 5 percent were reports of RCTs in 1976 (75). In an exhaustive search of the literature in English through 1981, Haines (103) found 51 RCTs related to neurosurgery; half of those had been published

BEST COPY AVAILABLE

since 1977. The growing number of RCTs and interest in them resulted in adding the heading "clinical trials" to *Index Medicus* in 1980. Another sign of this growing interest was the founding of the Society for Clinical Trials in 1978 to encourage exchange about methodological issues (see box A).

Box A.—The Society for Clinical Trials

The Society for Clinical Trials was founded in 1978 by a group of individuals with experience in clinical testing, epidemiology, statistics, and computer science. It was formed to allow greater exchange about methodological issues and about the impacts of RCTs, topics that are rarely addressed in medical periodicals, even in reports of trials.

The Society has more than 1,000 members. It sponsors an annual meeting and publishes the quarterly journal *Controlled Clinical Trials*. Its main objective is "to promote the development and exchange of information for design and conduct of clinical trials and research using similar methods." The society's specific long-term objectives include the following (209):

- Promotion of methodological research emphasizing design, organization, operation, and analysis.
- Promotion of the application of sound principles to design, operation through workshops and meetings sponsored by the organization. Some of these workshops and meetings may be international in character and held in countries other than the United States.
- Promotion of better communication by development, where possible, of standard terminology.
- Promotion of better understanding to those entering this field by serving as an important resource for the design and conduct for these studies.
- Promotion of better communication through the development of standards for the analysis and reporting of results.
- Promotion of better understanding by the general public of the importance of clinical trials for the evaluation of health care procedures.

A Description of the Method

General Structure

Fisher's rationale for randomizing as a valid basis for statistical inference is still the touchstone of RCT methodology. RCTs are actually a family of study designs that share the feature of randomized assignment to treatment groups.

In the simplest of these designs, individuals with a condition in common (e.g., the common cold) are allocated to two groups by an accepted randomization procedure (e.g., using random number tables or computer-generated random numbers). A promising but unproven technology (e.g., a new drug) is applied to one group, while the other is given the standard treatment, if one exists. The control group may be given no treatment at all, if that is standard, or preferably, when possible, a placebo that resembles the experimental drug. At an appropriate time after applying the technology each individual in the two groups is assessed for a prespecified outcome. The outcome can be death or a signal health event (e.g., a heart attack) or an intermediate physiological measure, such as a change in blood pressure. In a vaccine trial and some drug trials, presence or absence of disease after some time is an appropriate endpoint. The aggregate results for each group are then compared. Statistical tests are applied to the results to determine whether or not the new technology is better than the old.

In a well-designed trial, both the numbers of participants and the endpoints are chosen so that there is a reasonable probability that a statistically significant result can be obtained, if in fact the treatments being compared differ by some prespecified amount or more. While simple in theory, in practice RCTs are complex undertakings. Klimt (123) describes five phases in RCTs:

1. A Planning Phase that precedes general funding.
2. After approval of a broad outline of design and funding a Preparatory Phase, the protocol, the forms, and the organization are laid down in detail.
3. The Recruitment Phase that starts with the acquisition of the required number of clinical units and is followed by the recruitment of patients.
4. The Follow-up and Termination Phase during which no further recruitment takes place but patients are followed for the requisite number of years. The length of follow-up is determined by the nature of the disease and the kind of treatment effect expected. The termination part of this phase requires clean-up of the data base on patient information collected and final classification of endpoints.
5. Last, the Analysis Phase, where no new data are being gathered, the statistical analysis is performed, conclusions are drawn, and papers written.

Each phase presents its own challenges. The practical problems and basic guidance are discussed in the journal literature and in a limited number of texts, for example *Fundamentals of Clinical Trials,* by Friedman, Furberg, and DeMets (84), is an excellent reference. In addition, Peto and colleagues (180,181) provide a detailed description of RCTs for the nonstatistician, including both their design and analytic features.

The size and complexity of RCTs vary greatly. Small-scale pilot studies with only a handful of patients may be undertaken by a single researcher. At the other extreme, thousands of patients in centers around the world may be participants in a single trial. Many of the recent RCTs supported by the National Heart, Lung, and Blood Institute, particularly those in primary and secondary prevention of cardiovascular disease, are large multicenter endeavors. For example, the recently completed Multiple Risk Factor Intervention Trial randomized 12,866 men at 22 clinical centers to test the effect of a multifactor intervention program on mortality from coronary heart disease (166).

Although all well-designed RCTs require a great deal of effort and thought in design and execution, multicenter trials present greater practical problems. Well-conducted multicenter RCTs are characterized by such features as a centralized data collection center, a data monitoring committee (often of individuals independent of the study, with no vested interest in the trial or the intervention), and formal auditing procedures.

Blinding

Because of bias for or against a treatment on the part of researchers and patients, and to control for the effect of expectations of outcome, (a natural human characteristic), the element of "blinding" also has become a characteristic of RCTs. The object of blinding is to prevent the awareness of which treatment is administered. When only the patient is unaware of the treatment the study is "single-blind;" when both the person administering treatment and the patient are unaware, it is "double-blind." Additional layers of blinding can be added. Often a person other than the treating physician evaluates patient outcome. That person can in turn be unaware of which group a patient is in. The statistician analyzing the data may do so blinded.

The most valuable tool for achieving blinding is use of a placebo, an inactive substance or procedure that mimics the intervention tested, so that those who are to be kept blind cannot tell it from the active intervention. Placebos are most often used in drug trials, though at least one surgical RCT, assessing internal mammary artery ligation for coronary artery disease, used a sham operation as a placebo for the control group. That practice would not be acceptable today, since even a sham operation, involving anesthesia and operative incision involves risk. Ethical placebos can be developed for some procedures, however. A recent RCT of apheresis for schizophrenia used sham pheresis in the controls (see ch. 5). In some cases blinding is clearly impossible, as in comparing a surgical with a medical procedure, or when patients and physicians can identify a given treatment because of its special side effects. If blinding is not possible, the effect of bias in unblinded studies can be minimized to the extent outcomes are measured by objective standards. Whatever the outcomes measured, even with no blinding, randomized allocation will lead to more reliable results than any other type of allocation.

Techniques for Randomized Patient Allocation

Early randomization schemes were based on simple systems, such as the flip of a coin, alternate assignments of patients to groups as they arrived, or according to the day of the week they arrived, their birth dates, or their hospital or social security numbers.

Such methods have been abandoned for the most part, largely because the predictability of assignment allowed researchers and patients to manipulate assignments, or to selectively decide whether or not to participate in the trial. Assignments today are most often based on random number tables or computer-generated random numbers. Treatments may be assigned using presealed envelopes, opaque to the light. In multicenter trials, assignments are often computer-generated by a central office when a participant is enrolled, and given to the physician over the telephone, allowing little scope for physician bias in assigning treatments.

In theory, randomization of all individuals into requisite groups for a trial cannot be improved on. Given a large enough sample size, factors affecting outcome will be distributed more or less equally among the groups. Logically, for smaller numbers of people, randomization produces greater equality among groups the more homogeneous the population, and the fewer the prognostic factors that affect the outcome. In practice, because patients and resources are not unlimited, and often patient populations are rather heterogeneous, techniques have been developed to improve the distribution of the number of patients and their prognostic factors among groups.

The chance imbalance of numbers of individuals in the groups can be prevented by a special procedure called "random block permutation." In effect, this technique ensures that after some prespecified number of patients are entered in the trial, equal numbers are assigned to treatment groups.

"Stratification" is another commonly used, but controversial, method to better distribute factors of known prognostic importance during patient allocation. As individuals have entered the trial, they are classified by these factors, e.g., age, sex, and often diagnostic characteristics, e.g., extent of spread of a cancer. Randomization then takes place *within* these "strata," that is within these particular subgroups.

The value of stratification in patient allocation is not uniformly agreed on (137), but stratification in analysis is a generally accepted procedure. In the latter, adjustments are made after the data have been collected to adjust for chance imbalance in prognostic factors between groups.

"Minimization" is a more recent idea for patient allocation (218). The technique takes into account a number of variables of prognostic interest, up to 15 or more, without forming mutually exclusive subgroups. As each participant is entered, a series of calculations is made to determine which assignment would minimize the differences between the groups. Different weights can be assigned to different patient variables according to their prognostic importance. If all are given equal weight, group assignments are made simply to distribute equally the largest number of variables. Randomized allocation is used only in assigning the first patient and when there is a "tie" and the same difference between groups would occur regardless of assignment. Minimization has become popular particularly in cancer trials, where a large number of factors are known to have prognostic importance (184).

THE USES OF RCTs

The RCT was developed to discriminate between effective and ineffective treatments, particularly when the differences between treatments are moderate. More specifically, RCTs are used to accomplish the following:

- to compare the safety and efficacy of a new technology with a standard treatment, whether this is no treatment at all or a competing technology;
- to test the relative efficacy of a new technol-

ogy, assuming it has some other advantage over the standard, e.g., fewer side effects, lower cost;
- to determine the optimal way to use a technology to achieve a therapeutic effect; and
- to demonstrate the likely range of a technology's effectiveness in general practice as opposed to in highly controlled experimental settings. In a broader sense, RCTs can be used to answer questions susceptible to the scientific method about interventions involving human beings. Well-designed and executed RCTs are not merely product testing, but should answer questions about important hypotheses. They should, therefore, generate biologically and medically important information.

The results of RCTs may have widespread impact (143) insofar as they are used to allocate medical resources more efficiently (19,50,57,79,110, 143); to effect the adoption and use of medical innovations (70,89,91,113,143); to hasten the abandonment of ineffective therapies (11,111); and to resolve controversies about competing treatments (170).

RCTs are most useful when either the benefit of a new treatment is uncertain or the relative benefits of existing therapies are disputed (32). Thus, not all technologies need be evaluated in an RCT. Medical breakthroughs, such as the discovery of treatments like quinine for malaria, sulfa drugs and penicillin for bacterial infections, and insulin for diabetic acidosis, required no RCTs to demonstrate their efficacy. Startling breakthroughs, unfortunately, do not characterize most medical advances. Even in the case of breakthroughs, however, RCTs are useful to determine optimal treatment regimens. The current successful chemotherapy for Hodgkins disease was built up with stepwise RCTs after an initial breakthrough. Aside from breakthroughs, there are other technologies of accepted value that do not require the blessing of an RCT. For example (225):

> . . . cast application for forearm fracture is a technology whose efficacy has been established by experience in medical settings. It illustrates a technology whose efficacy could be called "manifest," that is, whose efficacy and safety are obvious to the observer. Although alternatives to cast application might be as efficacious, its widespread acceptance in this country makes development and testing of other methods unlikely and probably unnecessary.

THE ROLE OF THE PHYSICIAN IN RCTs

Traditionally, the physician has been the arbiter and judge of medical practices. It was presumed that careful observation of patients and reasoning about cause and effect would make the physician the best instrument to judge the success or failure of clinical practices. Until nearly the middle of this century, that presumption was largely unquestioned. Before the emergence of RCTs physicians were the only major actors in clinical decisionmaking. The growing importance of statistical evidence, and perhaps the growing importance of the statistician, was and is seen by some physicians as a threat. Some believe this response of physicians is a major impediment to the acceptance and adoption of good RCT results by the medical community (142):

> To some extent the clinician's marginalization was implicit in the rationale for the RCT. Not only was the RCT viewed as capable of making finer, more reliable discriminations between the relative merits of effective therapies (112), but randomization was introduced because of its superiority over the clinical investigator in controlling for the variables which might affect therapeutic outcomes. Moreover, early critics of randomization have noted, the goal of minimizing the investigator's interpretive role is implicit in the logic of statistical hypothesis testing.

The extent to which physicians' feelings of displacement have affected the development and impact of RCTs is impossible to assess. It can now be judged only by anecdotal evidence, precisely the standard that supporters of RCTs seek to replace. A more basic question than the one directly addressing RCTs may be a question about the role of research in general in clinical decisionmaking. Finally, it is important to understand the other factors that affect the way physicians treat patients.

Spodick (210) cites five behavioral pitfalls of physicians which affect both the conduct of RCTs and the acceptance of their results. The first is that the general acceptance of a practice is often taken for a proof of its effectiveness. The long use of bleeding, purging, and trephining provide examples. The rejection of "general acceptance" of a practice as adequate evidence of its efficacy underlay the 1962 amendments to the Food, Drug, and Cosmetics Act, which required "adequate and well-controlled studies" in support of new drug applications. Another pitfall of physician behavior is zeal, leading to glowing reports of success in the early applications of new practices. Such enthusiasm may be "inversely proportional to the quality of control" for treatments later shown ineffective or harmful in appropriately designed trials. Estrogen therapy for prostatic carcinoma, Vineberg implants for coronary artery disease, diethylstilbestrol to prevent spontaneous abortion, prophylatic portacaval shunts for portal hypertension, and internal mammary ligation are all practices that were enthusiastically embraced and have since been discarded because they lack efficacy or are unsafe.

A third pitfall is physicians' uncritical acceptance of poor data. Poor data are often given as much credence as good, and more if they support a preconceived notion of what is right. Often, because the sheer volume of poor data is so great, small amounts of good data are not visible. Long before diethylstilbestrol was known to be harmful to women who were exposed before birth, six well-controlled trials had shown that the drug was ineffective. Seven other uncontrolled or poorly controlled trials had taken precedence while 50,000 pregnant women per year took the drug. A fourth related pitfall is blindness to what data exist.

The final pitfall is the "it can't hurt mentality." Even when practices are proven ineffective through well-designed studies, they may still be continued. In some cases, no alternative treatment is available, and the physician feels that any treatment, even an ineffective one, is better than none. The physician may not always be wrong if "ineffective" is interpreted to include exploiting a placebo effect, or diverting patients from really harmful treatments. Unfortunately, however, there is never perfect knowledge about the effects of drugs or practices, and sometimes they may well "hurt" in the long term. The case of diethylstilbestrol illustrates this, as does the continued adherence to prescribing a bland diet, including cream, for peptic ulcer. There is some reason to believe that heavy intake of cream caused or accelerated atherosclerosis in some ulcer patients (40).

Spodick also speculates about the behavioral deterrents to initiating trials when they may be needed. Reverence for authority may cause physicians to adopt practices uncritically, i.e., when the practices are developed by and advocated by persons of renown. This was a factor in the widespread adoption of gastric freezing in treating peptic ulcer. Reverence for tradition makes it difficult to abandon an old practice, particularly when there is none to replace it. Physicians often feel a compulsion to treat, coupled with a reluctance to admit doubt. These attributes are often encouraged by patients. Physicians are also often loath to substitute clinical trial results for personal judgment in prescribing treatment. They may fear either withholding a new treatment or exposing patients to it, and therefore may be reluctant to participate in an RCT.

These views represent a fairly negative perception of physicians in relation to RCTs. On the positive side, it is physicians who initiate and participate in RCTs, and who form the majority of the method's proponents. As in most fields, acceptance of new methods is bound to be gradual, partly owing to appropriate skepticism. The use and impact of RCTs has grown since the 1940's, and the method itself is still evolving. Physicians and statisticians together are responsible for this progress, and there is evidence that physicians, including those in the community, are increasingly willing to participate in RCTs (see e.g., 65).

THE DIFFUSION OF MEDICAL TECHNOLOGY*

While it is useful to examine the effects of RCTs on the practice of medicine, it is useful to do so in the context of the larger questions of the adoption and use of medical technologies and the way medical practice changes.

The process by which a technology becomes part of the health care system is known as diffusion. Diffusion has two phases: the period when the decision is made to adopt the innovation, and the later period when decisions are made to use it. Research has focused on the first phase, as have Government policies. The use of a technology may be only tenuously related to its adoption. Each is discussed here in a separate section.

The Adoption of Technologies

The adoption of technological innovations has captured the attention of hundreds of researchers, resulting in thousands of articles and many theories (72). Early research grew out of sociology (192), but much recent work has been done by economists (195). A tacit assumption in much of this research is that adopting an innovation is desirable.

The classical model describing diffusion of technology is an S-shaped curve, based on the concept of "contagion" or "spread" (72). The diffusion of technologies such as intensive care units and cardiac pacemakers has followed this pattern (195,227). At least one other model, the "desperation-reaction model," has been described by Warner (246). A first phase of explosive diffusion occurs because of a provider's sense of responsibility to the patient and their mutual desperation faced with a life-threatening situation. These responses are related to what Fox (76) has called "scientific magic," which is partly the tendency of medical practitioners to favor vigorous treatments and to be staunchly hopeful even when a positive outcome is unlikely. Cancer therapies often fit the desperation-reaction model: there are few effective tools to fight the disease, and little time in which to act. In describing the model, Warner uses the example of chemotherapy for acute leukemia in children.

Before a technology is adopted or rejected it must be known. With regard to communication about technologies in the medical area, only the area of drugs has received the attention of researchers (120). Research on communication about drugs led to the description of a two-step model; information flows initially to physicians who are opinion leaders, and through informal channels, these leaders then transfer information to their followers (217).

The sources of information about technologies have been little studied. One study indicated that physicians specified drug companies' representatives as their most important source of information on new drugs (63). How the evaluations of technologies may affect their adoption has not been studied. It is clear, however, that the communication from researchers to practitioners is inadequate in both amount and quality.

A number of factors have been shown to influence the adoption of technologies. These include the characteristics of the technology, the complexity of understanding and using it, and the observability or visibility of its results (217). Characteristics of the adopter, including a cosmopolitan outlook have also been stressed (100). Large, complex, acute-care hospitals with medical school affiliations accept innovations more readily (176). Almost all the studies of adoption have focused on that of institutions like hospitals, and little is known about the adoption of technologies in practice situations.

Much research assumes physician dominance in decisionmaking (176). When there is concern about the slowness of change, physician conservatism is blamed. When premature adoption of technology is seen as the problem, physicians are considered to be uncritical and technology-hungry. Considerable homogeneity is assumed among physicians. Greer (101) has questioned these assumptions through research, still in progress, involving 362 focused interviews of those in the health care system, including 201 physicians. She found that community practitioners are generally not interested in gaining influence in the hospital, and have little effect on the acquisition of technologies. Medical technologies were more often acquired through the actions of hospital administrators and hospital-based physicians than at the demands of patient-admitting community physicians.

---

*This section is based on Banta, Burns, and Behney, 1982 (9).

From the standpoint of public policy, the key question is what characteristics of the medical environment affect adoption (96). These factors can be manipulated. They include financing methods, market conditions, and Government programs. The growth of third-party payment is without doubt related to the increasing use of medical technologies and increasing medical expenditures (167). The extent of coverage and methods of payment promote expensive hospital technologies and discourage preventive, rehabilitative, and ambulatory ones. Existing fee-for-service schedules reward the provider generously for diagnostic and curative services that rely on high technology. For example, a recent analysis in California showed that gastroscopy costs the physician $40 to $50, while Blue Shield pays up to $240 for the procedure (205).

A key regulatory program influencing adoption is the drug regulation program of the Food and Drug Administration (FDA). FDA is required to approve all new drugs as efficacious and safe before they are marketed. In 1976, FDA authority was extended to medical devices (see ch. 3 for a fuller discussion of FDA regulation). FDA processes generally slow the adoption of technologies. A considerable body of research has shown that the licensing of drugs in the United States is relatively slower than in other countries and that the lag can in part be attributed to FDA (200). Since many technologies have diffused prematurely, however, it is not clear whether this delay is good or bad. Many other Federal and State programs directly or indirectly affect the adoption of medical technologies through regulation and financial means.

The Use of Medical Technologies

While there are clearly some relations between adopting and using technology, they have not been clearly characterized. Some suggestive research in this regard has shown that hospital beds tend to be used regardless of the health problems or demographic characteristics of an area population (191). The ready availability of laboratory tests through automation has apparently stimulated their rapid increase (227). Cromwell and his colleagues, however, report that nonprofit hospitals in Massachusetts use certain diagnostic equipment at only 50 to 60 percent of capacity.

A surprising finding is the highly variable relation between patient needs and technology use (195). This is true even in the case of specific technologies addressed to clearly defined medical conditions. Wennberg and Gittelsohn (249) found that rates of common surgical procedures vary greatly in small areas of New England, for example, even when the areas are contiguous and demographically similar.

Physicians' training and their role in society are important factors in technology use. The sociological literature on professionalism and on physician dominance is large. Physicians are professionals granted a high degree of autonomy (80). They are also agents of the patient who attempt to provide the best possible care, regardless of cost. Because the patients pay little or nothing for procedures directly, and they work in a system that rewards the use of technology with both profits and prestige, physicians have strong reasons to use technology (247). The development of medical specialties has also affected technology greatly. Specialties have developed in response to professional, technological, and economic interests in the past (212), and will most likely continue to respond to these interests. The United States is faced with a potential excess of physicians (228), who could respond to the resulting pressure by entering specialty practice and maintaining their incomes by using specialized technologies more intensively.

Malpractice suits apparently encourage the use of technologies like skull X-rays (15), electronic fetal monitoring (8), Cesarean sections (140), and clinical laboratory testing (202). The dynamic nature of malpractice has been little studied. An overemphasis on technology and a correspondingly diminished concern on the part of the physician can dehumanize medical practice. Such dehumanization has been found to be associated with high rates of malpractice litigation (241).

Institutional factors affect technology use. The evaluations of prepaid group practices showing they led to fewer hospitalizations and less use of expensive technology were an important force in establishing health maintenance organizations in the 1970's. Similar evaluations are now encouraging the "competitive" strategy that is the latest policy. Such special medical institutions have been seen as a counterforce to the negative features of physician autonomy, but they also may diminish the physician's commitment to the interests of the patient and lead to a loss of the caring function of medicine (155). This could lead in turn to increased malpractice claims, a corresponding increase in technology use, and other problems.

As has already been stressed, fee-for-service payment to physicians and cost reimbursements to hospitals reward for providing more services. Existing fee scales reward more lucratively a physician's time using sophisticated technology than the physician's time in counseling (203). The spectacular rise in the use of ancillary services such as laboratory testing is related to specialization and extent of insurance, as well as payment methods. One study indicated that the greater use of nine such medical services accounted for about 40 percent of the increase in hospital operating costs from 1968 to 1971 (187).

The involvement of a profitmaking industry certainly affects the use of technology. The drug and device industries spend a large amount of money to promote their products. As mentioned previously, physicians say that the agents of drug firms are their most important sources of information about drugs.

Abandonment of Medical Technologies

While researchers have been enthralled with the adoption of technologies, little has been done toward understanding their abandonment. McKinlay (150) decribes a commonsense view of the "erosion and discreditation" of medical technologies. The initial enthusiasm for the technology when it was an innovation wanes and its applications are not so global as once thought. Sometimes a scandal abruptly cuts short the life of a technology, thalidomide, for instance. More often, it is eclipsed by a new innovation. Finally, McKinlay says, "it is relegated to that great dust heap called History."

In one of the very few attempts to analyze the abandonment process using empirical evidence, Finkelstein and Gilbert (72) examined the decline in use of eight drugs over the period 1964 to 1982. Seven had been introduced between 1963 and 1972, after the 1962 Amendments to the Food, Drug, and Cosmetic Act (see ch. 2) and one, tolbutamide (a hypoglycemic agent used by diabetics to lower blood sugar) which had been introduced earlier, but which experienced its decline during the later period.

Finkelstein and Gilbert began, for the sake of argument, with the assumption that abandonment would share features with adoption: that opinion leaders would first act on negative information about a drug, followed by the rest of the medical profession. Such a pattern represents the S-shaped curve. Their results suggest that, for the eight drugs studied, the pattern of abandonment does not fit the S-shaped curve. Declines in use were generally more precipitous, arguing that perhaps "physicians are sometimes affected directly by external information stimuli without the need for processing by an intermediary opinion leader." Based on their findings, Finkelstein and Gilbert suggest that more investigations using empirical data could profitably be undertaken to systematically characterize alternative models for the abandonment and adoption of medical technology. The ultimate value might lie in better understanding of the influences on physicians in adopting and abandoning technologies.

RCTs and the Diffusion Process

As the preceding sections have indicated, the reasons that medical technologies are adopted and used are far more complex than "simply" evaluating the evidence from RCTs and making reasonable decisions on that basis. The impacts of RCTs must be seen in this broader context, and efforts to increase their impact must consider the economic, regulatory, and institutional influences on adopting and using medical technologies.

The appearance of RCT results is not the start of a decisionmaking process about a medical practice, but comes after some diffusion has already taken place. Physicians may already have some personal experience with the technology, which may sway them in one direction or the other. RCTs are rarely conducted before new technologies are widely diffused (201). Banta and Thacker (8) document the widespread diffusion of electronic fetal monitoring despite the lack of evidence that it improves birth outcomes.

RCTs figure in two distinct processes: synthesis and consensus development. Synthesis is the process of integrating the findings from different studies and developing generalizations based on the results. All types of studies, both laboratory and clinical, may be considered in synthesis. Techniques for synthesis range from elementary qualitative procedures to sophisticated statistical manipulations.

The traditional approach to synthesizing research is the literature review. Typically, a reviewer selects a set of studies believed to be most relevant and summarizes the evidence. Because of the limitations inherent in literature reviews, efforts have been made to develop more systematic procedures to integrate and interpret sets of research evidence.

A simple structured synthesis technique involves organizing a body of literature according to a prespecified set of criteria and is actually a classification procedure (135). Sometimes called the "voting method," this synthesis technique involves selecting a particular sample of evaluative studies of a technology, coding some aspect of the design and/or conceptual framework, classifying observed outcomes as to whether they are favorable, neutral, or unfavorable (i.e., "taking a vote"), and then constructing tables of research findings.

A rigorous statistical approach to research synthesis is a quantitative synthesis technique called meta-analysis (93). This technique uses the actual results of studies and permits the determination, across a set of studies, of the magnitude of treatment impact. Meta-analyses are useful in assessing treatments for which a large number of studies are available and findings across studies seem to have great variability.

A number of organizations carry out synthesis activities. OTA reports have included a number of syntheses of specific technologies. Case studies prepared for *The Implications of Cost-Effectiveness Analysis of Medical Technology* (229) synthesize results of all types of research in their assessments. The activities of the former National Center for Health Care Technology and currently the Office of Health Technology Assessment (National Center for Health Services Research, Department of Health and Human Services), are synthesis activities carried out by the Federal Government, in general with the aim of making statements about risks and benefits of technologies. In the private sector, the American College of Physicians and the Blue Cross and Blue Shield Association have specific programs of medical technology evaluation which use synthesis techniques.

Consensus development is a group decision process designed to produce a "consensus statement" about a medical technology, that can be accepted by clinicians, researchers, and the public. The statement should identify what is known and not known about the technology, in terms of the safety, efficacy, and appropriate conditions for use. The major sponsor for consensus development is the National Institutes of Health, through the Office of Medical Applications of Research. Unlike some of the structured synthesis techniques, consensus development conferences have no specific theoretical basis for their format. Consensus statements are widely distributed by NIH to the leading medical journals.

2.
Funding for Clinical Trials

2.
Funding for Clinical Trials

Clinical trials, like all good research, can be expensive. The more participants they engage and the longer the trial runs, the more expensive they become. Two recent multicenter randomized clinical trials (RCTs) sponsored by the National Heart, Lung, and Blood Institute (NHLBI) are budgeted at over $100 million (240). While these are the most expensive trials undertaken in this country, costs over $1 million are common. The cost of clinical trials is one of the main factors driving the search for alternate methods to answer the same questions. Nevertheless, RCTs are now the superior means to evaluate the efficacy of medical technologies. Insofar as RCTs contribute to more rational decisionmaking in health care, halting the adoption and hastening the abandonment of ineffective technologies, their immediate costs may be justified. Nonetheless, only a limited number of trials can be funded. At the moment, the funding of clinical trials is concentrated in biomedical research and drug development programs. In the coming years, however, judging from current discussion related to funding, their costs may be more widely spread throughout the health care system.

A large number of clinical trials in this country are supported by the Federal Government. The drug industry is also a major supporter of trials of proprietary drugs, the results of which are used to gain approval of new drugs by the Food and Drug Administration (FDA) (table 1). Many other private health and medical groups, such as the American Cancer Society and the American Heart Association, fund a small number of trials, but generally these are not the large-scale, multicenter trials like those that the Federal Government or industry can support.

In 1979, the companies represented by the Pharmaceutical Manufacturers Association (PMA) (over 90 percent of companies in the industry) spent about $212 million on clinical evaluation, a figure including phase I, II, and III clinical trials (182). RCTs are generally conducted in phase III, but no more detailed breakdown of expenditures for RCTs is available from PMA. In any case, it is a substantial sum of money.

Table 1.—Studies Required in FDA's Premarketing Drug Approval Process

Phase I:
- Studies in normal volunteers or relatively healthy patients to determine safety and pharmacologic effects.
- Small studies in patients to determine clinical effectiveness.
- Total number of subjects—up to 80 administered the investigational drug.

Phase II:
- Controlled clinical trials to determine appropriate doses, safety, and effectiveness.
- Total number of patients—about 200 administered the investigational drug.

Phase III:
- Controlled and uncontrolled clinical trials to determine safety and effectiveness and to support labeling claims.
- Total number of patients—about 500 to 3,000 administered the investigational drug.

SOURCE: U.S. Food and Drug Administration.

The largest supporter of clinical trials in the Federal Government is the National Institutes of Health (NIH). The Alcohol, Drug Abuse, and Mental Health Administration (ADAMHA) finances RCTs under its Treatment Assessment Research Program. The Veterans Administration (VA) supports multicenter RCTs in VA medical centers through the Cooperative Studies Program.

The U.S. Department of Defense (DOD), largely through the Department of the Army, supports a large field studies program, conducting RCTs mainly of vaccines and prophylactic drugs and of some treatments.

Academic institutions also support RCTs, mainly in the form of researchers' salaries. The dollar value of this contribution is not known (158).

Of equal interest is who does not fund clinical trials. Third-party payers for medical care generally do not. Because clinical trials, and RCTs in particular, are important in assessing technologies and in better decisionmaking, they should be of great interest and value to third-party payers. The accelerating costs of health care have led to concern of third-party payers about costs and about covering only those medical practices of proven value. The RCT is the best method for gathering evidence on the effectiveness of a practice, in cases where the method is appropriate.

The greatest expense in conducting RCTs is patient care. At present, the VA system excludes from the research budget nearly all patient care costs in RCTs. Under most other funding arrangements, research money covers varying percentages of patient care and institutional (hospital) costs in RCTs as well as the associated costs of trials. The research community is now active in encouraging private third-party payers to increase their contributions to patient care costs in RCTs.

Through the Medicare program, the Federal Government directly pays about one-quarter of all third-party medical payments in the United States. The large and ever-rising cost of health care, symptomatic of today, is a powerful incentive toward more informed decisionmaking. Congress, in the Social Security Act Amendments of 1983, recognized the need for reliable assessments of medical technologies by, for the first time, allowing the Health Care Financing Administration (HCFA) to fund RCTs relevant to their needs for information (for a fuller discussion of policy decisions under HCFA, see ch. 3). HCFA's will undoubtedly be an important contribution to RCT financial support.

TRENDS IN FUNDING CLINICAL TRIALS

Trends in Federal funding of clinical trials were encouraging through the 1970's. Between 1971 and 1974, 4 of the 11 NIH institutes—NHLBI, the National Cancer Institute (NCI), the National Institute of Neurological and Communicative Disorders and Stroke (NINCDS), and the National Eye Institute (NEI)—nearly tripled their obligations for major clinical trials, including RCTs (225). In 1979, NIH expenditures for clinical trials totaled $136.2 million, in support of 986 trials. The numbers have increased steadily since 1975 when $87.8 million went to support 755 trials. The amount spent on clinical trials as a percent of total NIH expenditures, 4.3 percent in 1979, has changed relatively little during that time, however. Since 1979, comparable data have not been compiled, but evidence suggests a downturn in the support of clinical trials, brought about by budgetary constraints and policies concerning the total number of competing grant awards (235). In the *National Institutes of Health Research Plan, FY 1983-85,* NHLBI states (239):

> . . . the most severe impact [of holding the number of grants constant] will be felt in clinical trials and targeted research, funded under the contract mechanism, where no new efforts can be implemented in 1980-1982. . . . The contract mechanism is best suited to fund clinical trials, and rapid advances in research and developments in cardiovascular and pulmonary treatment techniques necessitate clinical evaluation at a time when no new contracts can be awarded.

Some of the other institutes make similar statements (235).

Funding for VA's multicenter clinical trials increased throughout the 1970's. In fiscal year 1970, VA spent $1.8 million, 3.1 percent of its total budget for biomedical research and development, on clinical trials. By 1981, the figure was $9.7 million, representing 7.1 percent of this VA budget.

THE NATIONAL INSTITUTES OF HEALTH

The major biomedical research agency in the United States, NIH, is also the largest supporter of clinical trials.

Clinical trials included in NIH statistics cover more than just RCTs. According to the *NIH Inventory of Clinical Trials,* clinical trials are defined as follows (242):

> . . . a scientific research activity undertaken to define prospectively the effect and value of prophylactic/diagnostic/therapeutic agents, devices, regimens, procedures, etc., applied to human subjects. It is essential that the study be prospective, and that the number of cases or patients will depend on the hypothesis being tested, but must be sufficient to permit a definite result to be anticipated. Phase I, feasibility, or pilot studies are excluded.

Of NIH trials active in 1979, about 60 percent were RCTs (158), up from about 50 percent in 1975 (225).

The emphasis given to clinical trials varies considerably from institute to institute. NCI and NHLBI, the largest institutes, are also the largest supporters of clinical trials (table 2). These NIH institutes differ somewhat from the others as they are the only ones specifically mandated by acts of Congress, and clinical research is specifically mentioned in their legislation. The other institutes are guided by the general research authority of the Public Health Service Act, which provides a less specific mandate (235).

NIH institutes least active in clinical trials are the National Institute for Environmental Health Sciences (NIEHS), which supported no clinical trials in 1979, and the National Institute of General Medical Sciences (NIGMS), which supported one. NIEHS is mainly concerned with the adverse effects of environmental factors on human health. Such effects are not readily studied in clinical trials. NIGMS primarily supports undifferentiated basic research, that does not necessarily focus on a specific disease. Technologies ripe for clinical trials are usually no longer in the purview of NIGMS.

The seven remaining institutes fall between the two extremes, their use of clinical trials dictated to some degree by the state of knowledge of the diseases they study, and to a large extent by the importance accorded clinical trials by key individuals within the individual institutes. The National Institute of Arthritis, Diabetes, and Digestive and Kidney Diseases (NIADDK), for example, supports a great deal of clinical research on the mechanisms of the chronic diseases. NIADDK is now testing some promising treatments for these diseases (e.g., apheresis for a number of conditions), but there are not within its purview at this time as many promising technologies ready for clinical trials as there are, for instance, in the areas of heart disease and cancer. NCI has strongly supported RCTs since the late 1940's, even before very many promising cancer treatments had been developed. It was farsighted statisticians and other researchers working in the cancer field that provided the impetus. NEI supported no RCTs 15 years ago; it now supports more than 20, stimulated in large part by a few motivated advocates (see box B).

In the mid-1970's, NIH began to compile an annual inventory of the clinical trials supported by all its institutes. Data were collected in 1974, and the first published compilation covered trials active in fiscal year 1975. The last compilation was of trials active in fiscal year 1979. NIH no longer compiles these data on clinical trials. Some but not all of its institutes have continued inventories for their own purposes, in the same form as they did for the NIH-wide inventory. NCI publishes a *Compilation of Experimental Cancer Therapy Protocol Summaries,* which includes phase I, II, and III studies, a much broader range of trials than were included in the NIH inventory.

NIH inventories summarized data from each trial on a standard survey form. Clinical trials were defined to include more than RCTs, but to exclude very small trials, phase I drug studies, and feasibility and pilot studies. The data collected described the trials purpose, starting date, type and amount of support, subject population, administration, and other characteristics. The summaries classified trials by type and amount of support, number of participants, type of experimental design (e.g., randomized or nonrandomized assign-

Table 2.—NIH Support for Clinical Trials, Fiscal Year 1979

A.—Amount of NIH Support for Clinical Trials Active in Fiscal Year 1979, by Institute for Type of Support

Institute	Extramural support				Intramural support[b]	Total amount of support
	Grant	Contract[a]	Grant and contract	Total		
NIH	$47,304,588[c]	$75,738,768	$1,954,960	$124,998,316	$11,161,800	$136,160,116[c]
NEI	3,141,547	5,378,262	—	8,519,809	85,800	8,605,609
NHLBI	4,006,736	50,933,477	159,788	55,100,001	1,423,500	56,523,501
NIAID	2,435,341	3,827,597	—	6,262,938	234,000	6,496,938
NIAMDD	1,927,658	5,226,975	—	7,154,633	1,085,500	8,240,133
NICHD	3,074,448	556,296	—	3,630,744	552,500	4,183,244
NIDR	221,977	557,672	—	779,649	999,050	1,778,699
NINCDS	1,786,449	439,000	—	2,225,449	435,500	2,660,949
NIGMS	225,750	—	—	225,750	—	225,750
NCI	30,484,682[c]	8,819,489	1,795,172	41,099,343	6,345,950	47,445,293[c]

[a] Contract includes interagency agreements without intramural support.
[b] Intramural support includes intramural support in combination with interagency agreements.
[c] One trial did not report amount of support.

SOURCE: National Institutes of Health, *1979 Inventory of Clinical Trials.*

B.—Number of Clinical Trials Supported by NIH in Fiscal Year 1979, by Institute for Type of Support

Institute	Number of trials supported extramurally				Number of trials conducted-intramurally[b]	Total number of trials
	Grant	Contract[a]	Grant and contract	Total		
NIH	592	212	11	815	171	986
NEI	20	3	—	23	3	26
NHLBI	3	13	1	17	3	20
NIAID	80	34	—	114	6	120
NIAMDD	30	22	—	52	15	67
NICHD	24	6	—	30	2	32
NIDR	2	11	—	13	13	26
NINCDS	17	3	—	20	20	40
NIGMS	1	—	—	1	—	1
NCI	415	120	10	545	109	654

[a] Contract includes interagency agreements without intramural support. Two trials were supported mostly by contract with some intramural support.
[b] Intramural support includes intramural support in combination with interagency agreements. One trial was supported mostly by intramural support with some contract support.

SOURCE: National Institutes of Health, *1979 Inventory of Clinical Trials.*

ment of participants to groups, use or lack of a control group, type of control group) and type of intervention (e.g., therapeutic, diagnostic, or prophylactic).

The NIH inventory was managed by the Division of Research Grants which, for the first 2 years, supported it with funds designated for evaluation. As resources and personnel became scarcer, funding the inventory became increasingly difficult. Collecting the information itself was not easy, although the institutes experienced different degrees of difficulty in providing the needed information. The future of the inventory is unclear, but without some measure like the inventory, trends in clinical trials are hard to document.

In 1979, total NIH clinical trials of therapeutic interventions, 494, far outnumbered those of prophylactic interventions, 118, and diagnostic ones, 53 (table 3). Among the 1979 trials, however, prophylactic trials cost most, $59 million, compared with the $51 million NIH spent on therapeutic trials and the $3 million it spent on diagnostic ones. The discrepancy in order between the two sets of figures arises because the large-scale multicenter prevention trials funded by NHLBI, while few in number, are relatively expensive. In

Box B.—The National Eye Institute

Soon after the National Eye Institute (NEI) was established in 1968, it began the Diabetic Retinopathy Study (DRS). First recommended by the Advisory Council to the then National Institute of Neurological Diseases and Blindness, the study assessed laser treatment used to halt the progress of vision loss in patients with proliterative diabetic retinopathy. Retinopathy, one of the major complications of insulin-dependent diabetes, is a leading cause of blindness in this country. Assessing such laser therapy by an RCT was extremely important.

The significance to ophthalmology of DRS is even greater, marking the beginning of a trend in the field's clinical research. Since the mid-1950's, when RCTs confirmed that high dosages of oxygen to infants in incubators caused retrolental fibroplasia, no major RCTs had been carried out in ophthalmology in this country. DRS established the use of RCTs in the field. NEI shortly thereafter funded two more large RCTs under contract, one a direct successor to DRS. NEI now funds more than 20 RCTs, most grant-supported.

There are some readily apparent reasons for the success of RCTs at NEI, many of them related to DRS. The first and present Director of NEI, Carl Kupfer gave high priority to clinical trials generally, and believed it part of NEI's mission to carry out RCTs. He established the Office of Biometry and Epidemiology to manage contract-supported RCTs, which became a national focal point for RCTs in eye disease.

The DRS was well designed and well run. It had an unequivocally positive outcome: Laser treatment did prevent blindness by almost 50 percent over the 5-year period of the study. Finally, it involved a large number of ophthalmologists in 15 clinical centers. Participating in or knowing about the study sensitized ophthalmologists to RCT methods. This accounts, to some degree, for the increased number of NEI grant applications for RCTs.

In addition to supporting RCTs, for nearly a decade NEI has taught an annual short course on clinical research methods at the American Academy of Ophthalmology.

Table 3.—Number and Amount of Support for NIH Supported Clinical Trials Active in Fiscal Year 1979, by Institute for Type of Intervention

	Total trials supported in fiscal year 1979[a]		Type of intervention					
			Therapeutic[a]		Prophylactic[a]		Diagnostic[a]	
Institute	Number[b]	Amount[b]	Number	Amount	Number	Amount	Number	Amount
NIH	666	$112,847,367	494	$50,540,964	118	$58,875,778	53	$3,170,625
NEI	26	8,605,609	22	4,890,194	2	3,415,997	2	299,418
NHLBI	20	56,523,501	10	9,726,605	10	46,796,896	—	—
NIAID	120	6,496,938	57	2,992,347	39	2,697,064	24	807,527
NIAMDD	67	8,240,133	60	7,680,072	4	246,798	3	313,263
NICHD	32	4,183,244	16	2,532,054	15	1,629,175	1	22,015
NIDR	26	1,778,699	7	779,051	17	776,871	2	222,777
NINCDS	40	2,660,949	35	1,565,020	2	959,429	3	136,500
NIGMS	1	225,750	—	—	1	225,750	—	—
NCI	334	24,132,544	287	20,375,621	28	2,127,798	18	1,369.125

[a] Trials in cooperative groups not included.

[b] One trial did not report amount of support. One trial did not specify type of intervention.

SOURCE: National Institutes of Health, *1979 Inventory of Clinical Trials*.

BEST COPY AVAILABLE

1979, the average cost of a clinical trial at NHLBI was about $2.8 million, the highest average cost of all the institutes. The National Institute of Allergy and Infectious Diseases (NIAID), for example, spent an average of $54,000 per trial.

Most NIH-sponsored clinical trials are conducted extramurally. In 1979, of all 986 NIH trials, only 171 were conducted intramurally (by a scientist on the NIH campus). Extramural trials are funded through either grants or contracts, with the mix in types of funding varying among institutes. Overall, they spend about twice as much on contracts as on grants, although this statistic may disproportionately reflect the pattern of one large institute, NHLBI. The institutes together fund a larger number of trials by grants (592 v. 212) though again this reflects the large number of smaller trials funded by one institute, NCI. Larger, multicenter trials are probably more appropriately funded under contracts, which presumably give the sponsoring institute more control over the trial, while small, single institution trials probably are more appropriately funded by grants.

NIH has greatly fostered the use and development of RCTs from the early work in cancer chemotherapy, to the large-scale trials in heart disease. These trials have contributed not only to that body of knowledge of medical practices derived from testing with RCTs, but also to the improvement and sophistication of the RCT method itself. Specific trials and groups of trials of particular medical significance are discussed in chapter 5.

THE ALCOHOL, DRUG ABUSE, AND MENTAL HEALTH ADMINISTRATION

ADAMHA, an agency of the U.S. Department of Health and Human Services, is composed of three institutes, each devoted to programs of basic and applied research, service, and training, in its own area: the National Institute on Alcohol Abuse and Alcoholism (NIAAA), the National Institute on Drug Abuse (NIDA), and the National Institute of Mental Health (NIMH). ADAMHA and its predecessor agencies have conducted research to establish the safety and efficacy of medical technologies since the 1950's. In 1975, however, ADAMHA established Treatment Assessment Research (TAR) as a separate kind of research, to study the relative safety and efficacy of various therapeutic substances and procedures applied to human subjects. This research includes clinical trials, case reports, retrospective surveys, and reanalysis of early data (225). In 1982, the total TAR budget was $18.5 million (125). Table 4 gives a breakdown of expenditures by institute. The amount spent specifically on RCTs is not available. Of the three institutes, however, NIMH most actively promotes clinical trials (see box F in ch. 5).

Table 4.—ADAMHA Treatment Assessment Research Fiscal Year 1982 Expenditures

Institute	Millions of dollars
National Institute of Mental Health	$12.778
National Institute on Drug Abuse	4.995
National Institute on Alcohol Abuse and Alcoholism	.774
Total	$18.547

SOURCE: R. Kopanda, ADAMHA, personal communication.

THE VETERANS ADMINISTRATION

The VA Cooperative Studies Program (CSP) supports multicenter clinical trials within the VA medical care system. As of September 1982, CSP had 19 studies in the implementation stage (all but 2 RCTs), 11 in active planning, and 12 in final analysis. The technologies the VA studies reflect the medical problems of the veteran population. Of ongoing and recently completed studies, the greatest number treat cardiovascular disease. VA research also emphasizes alcohol-related diseases, and dental and mental conditions. Other VA trials treat acute infectious diseases, diabetes, epilepsy, and conditions associated with disabling injuries. The largest number of trials have tested drug therapies, followed by those that have tested types of surgery. While most trials have concerned treatments, many have focused on the prevention of cardiovascular disease through control of hypertension. The mix of VA clinical trials is much like that of NIH, except that fewer VA trials focus on cancer treatment.

CSP is centrally administered at VA headquarters in Washington, D.C., and has four centers to coordinate data and one experimental drug unit located in different parts of the country.

CSP trials follow a well-defined pathway from inception to final analysis and publication. Ideas for studies come from physicians and investigators in VA installations around the country. They are considered by VA panels and outside advisors, and if judged appropriate for VA research and worthwhile are planned and carried out. Each study is assigned a coordinating center for help in design and conduct of the trial including final data analysis. This procedure ensures the high quality of the study's design and implementation, and obviates the need that the principal investigator be an epidemiologist or statistician.

Up to the present, all the ideas for VA studies have flowed from the "provinces" to the central office. The CSP office in Washington is now beginning to encourage studies that are important, as well as continuing to receive ideas from the field.

The deceptively small budget of CSP, about $12 million per year, goes mainly to support the coordinating centers and other nontreatment aspects of the trials. In contrast to the funding procedure for clinical trials through other mechanisms in this country, in VA trials the participants' treatment in trials is paid for *entirely* through a different channel, in this case, as VA medical benefits.

CSP only supports trials that require the participation of more than one VA hospital. Other clinical trials are conducted within single VA hospitals, and VA is involved in trials funded by other sources (e.g., NIH, pharmaceutical companies), but there is no central register of these activities.

THE DEPARTMENT OF DEFENSE

DOD is a major supplier of health care in this country. Many of the health problems it must treat, however, differ from those of the civilian population. DOD also conducts much health-related research, most of it directed toward developing and testing drugs and, especially, vaccines. A significant part of this research is conducted entirely by DOD, particularly by the Department of the Army, from drug and vaccine development all the way through large-scale field testing in RCTs.

The Department of the Army is now conducting between 60 and 70 drug and vaccine studies in humans, including studies in phases I, II, and III. RCTs are now under way on a vaccine for gonorrhea, the use of steroids in life-threatening typhoid, antileishmania agents, and antibiotic prophylaxis of leptospirosis. For the past 20 years, the Army has supported a development program for antimalarial drugs that relies heavily on RCTs for final recommendations on prophylaxis and treatment. These recommendations form the basis for practice worldwide (34).

Results of DOD vaccine trials and some drug trials have provided information for DOD policymaking, and DOD's recommendations are often adopted by the civilian population. Among the vaccines developed and tested wholly or in part by DOD are those for meningococcus, adenovirus, typhoid, yellow fever, Rift Valley fever, Venezuelan encephalitis, Rocky Mountain Spotted fever (now in late stages of field testing), and gonorrhea (soon to be tested). DOD is also involved in national efforts to develop influenza vaccines. All of its modern vaccine developments have included large-scale field testing in RCTs.

DOD has no central mechanism to track RCTs in its system. In theory, individuals at any DOD installation can carry out trials, and each branch of service is autonomous in conducting RCTs, unless the cooperation of other branches is required, e.g., for trials that require large subject populations. DOD has no regular coordinating body or mechanism to facilitate multicenter or multibranch trials. Each trial is done ad hoc. The Department of the Army keeps most of its financial information on RCTs by subject area (e.g., malaria, typhoid, etc.), so the total amount of money it spends on clinical trials is not easily compiled.

HEALTH INSURERS AND SUPPORT OF RCTs

A growing recognition of the value of RCTs in making sound coverage decisions by both public and private third-party health insurers has manifested itself recently in a number of ways, and has brought several basic issues to the fore. A central issue is to define the appropriate role for third-party payers in supporting RCTs. It is probably unrealistic to expect insurers to underwrite RCTs entirely. A more reasonable expectation is that they cover a greater share of the costs of treating trial participants. Currently, a prohibition against paying for experimental or investigative procedures exists in most private health insurance contracts. Insurers do reimburse for some patients in RCTs receiving "standard" care. This might mean patients in control groups, or even patients in "experimental" groups if the RCT is evaluating a practice already in use. RCTs often require more lab tests and closer observation of all patients, experimental and control, than a patient would receive under nontrial conditions. These excess costs, which may be substantial, are not generally covered by third-party payers.

A more significant reason for lack of sponsorship of RCTs by private health insurers is the administrative structure of those companies. The Blue Cross and Blue Shield Association, the largest private health insurer, is not capable of requiring that individual plans (State and local) and individually insured groups contribute to clinical trials in general or particular trials. This is because each group that seeks health insurance through local Blue Cross or Blue Shield Plans contracts for coverage for that group alone. Some of these groups may be as small as 50 enrollees while others are national accounts with hundreds of thousands of employees spread across several States (169).

In one of the few examples of third-party reimbursement for both the study treatment and sham treatment, five State and local Blue Cross/Blue Shield groups and other third-party payers agreed to reimburse five centers involved in an RCT of plasmapheresis v. sham pheresis for multiple sclerosis. HCFA and the State Medicaid groups, on the other hand, are not participating. Thus, patients' eligibility for the trial depended not only on medical criteria, but also on the type of health insurance they had. The administrative and other research costs of the trial are funded through an NIH grant. While the trial is successfully under way, getting agreement from the third-party payers was a cumbersome and time-consuming process.

In another example, all funds for patient care are being provided by third-party payers in a trial of "extracranial/intracranial bypass," a surgical procedure to prevent stroke in patients with cerebrovascular disease. This multicenter study involves 20 major medical centers in this country and three outside the United States (147). The National Institute of Neurological and Communicative Disorders and Stroke (NINCDS) is supporting the administrative costs of the central office and the data center, and the costs of entering and following up patients. Hospitalization and medical fees are covered by the third parties (97).

Some of the current activities concerning third-party payers and RCTs are described below. The Institute of Medicine of the National Academy of Sciences is considering the role of third-party payers in clinical trials as one aspect of its project on "Evaluating Medical Technologies in Clinical Use."

The Arthritis Foundation and the National Multiple Sclerosis Society are sponsoring a meeting (to be held in July 1983), at which they hope to develop a proposal for the participation of third-party payers in funding clinical trials. Representatives of the private insurers as well as the Government will attend the meeting.

3.

RCTs and Health Policy

3.

RCTs and Health Policy

Randomized clinical trials (RCTs) play a direct role in one major area of health policy: the regulation of drugs and, to a lesser extent, of medical devices, both by the Food and Drug Administration (FDA). FDA requires that for all new drugs, and for certain devices, evidence of safety and efficacy must be shown before they are approved for marketing. The standard of evidence is the RCT. In other health policy areas, RCTs figure less prominently. No Federal agencies directly regulate medical practice, and no governmental body requires proof that medical practices are safe and effective before they can be used. Institutional review boards of individual medical institutions are responsible for ensuring that research projects meet ethical standards. There are no legal constraints and there may be no institutional constraints to introducing new procedures not labeled as research.

The other major area in which RCTs can affect medical policy is in decisions about payment for medical practices by health insurers. Since most medical practices have not been assessed by RCTs, it would be unrealistic to expect health insurers to cover only the practices that have been. In fact, until perhaps a decade ago, third-party payers usually accepted uncritically the judgment of physicians about what was appropriate patient care, and reimbursed on that basis. The rising costs of health care, in large part attributable to the rise of high-technology medicine, have forced insurers to look more closely at what they are paying for. The Federal Government, the largest third-party payer in the country through Medicare, has a stake in ensuring that the health care it pays for is "reasonable and necessary," as statute dictates. Though RCT results have been available for few coverage decisions so far, the potential for their use in decisionmaking by the Government and private third-party payers is substantial.

Private health insurers and health maintenance organizations generally have more latitude in coverage decisions than the Federal Government since the coverage they provide is not a matter of law, though it is a matter of contract. The benefits packages each insurer offers may be different, to appeal to different clientele. An even greater role for RCTs can be envisioned in those circumstances where decisions about medical practices could be made based on cost-effectiveness criteria rather than on the more inclusive criteria of "reasonable and necessary." Blue Cross/Blue Shield, the largest private insurer, has begun to look at medical practices through their "Medical Necessity Project," which began as an attempt to identify obsolete practices, and has evolved into a mechanism for making decisions about coverage of new and existing technologies. RCTs should thus be of greater and greater importance for private insurers as the most reliable source of information about the efficacy and safety of medical practices.

De facto regulation of medical practice by third-party payers through coverage and reimbursement decisions will probably never become as regimented as, for example, the drug approval process. Such regimentation would be stifling to medical practice and a threat to innovation. The goal of responsible regulation in this is not to attain uniformity of medical practice, but to assure that decisions be made with the best information, including—when appropriate—the results of RCTs.

DRUG REGULATION

The approval of new drugs in this country provides an unambiguous role for RCTs in policymaking. By statute, new drug approval requires the submission to FDA of the following:

> . . . "substantial evidence" . . . consisting of adequate and well-controlled investigations, including clinical investigations, by experts qualified by scientific training and experience to evaluate the

> effectiveness of the drug involved, on the basis of which it could fairly and responsibly be concluded by such experts that the drug will have the effect it purports or is represented to have under the conditions of use prescribed, recommended or suggested in the labeling or proposed labeling thereof (sec. 355(d)).

The section of the Food, Drug, and Cosmetics Act requiring "substantial evidence" is part of the 1962 amendments to the original 1938 act. The 1938 legislation for the first time required that drugs be "safe," but did not require any evidence of their effectiveness. Decisions about drug effectiveness were left to the clinical judgment of physicians. RCT methodology was still developing, and the method was little used at that time.

The precipitating factor behind the 1962 amendments was a drug-related disaster. Alarm arose with the recognition that thalidomide, a tranquilizer, caused grossly abnormal limbs (phocomelia) in babies of women who had taken the drug while pregnant. Thalidomide was available in Europe, but had not, in fact, been approved in this country. People obtained the drug in this country under Investigational New Drug protocols or by purchasing it abroad.

The problem of thalidomide was not efficacy. (Thalidomide was an effective tranquilizer.) What emerged in the amendments as a result of the thalidomide case, however, was the requirement that new drugs be effective as well as safe. The history of the substance of the amendments is anything but straightforward. Most of it is unrelated to drug efficacy or RCTs, and it will not be discussed here in detail. (For a brief history of drug regulation and the new drug approval process, see ref. 171.)

The authors of the 1962 amendments were not necessarily thinking of RCTs when they wrote the phrase "adequate and well-controlled studies." That language may simply have been obtained from testimony in hearings. The phrase was used as the scientific analog of the legal phrase "substantial evidence" (i.e., more than an iota, less than a preponderance).

The details of what constitutes adequate and well-controlled studies were published in FDA regulations. The section "refusal to approve the application" (314.111) lays out the kinds of evidence required for drug approval. The Commissioner may refuse to approve the application when:

> (5)(i) Evaluated on the basis of information submitted as part of the application and any other information before the Food and Drug Administration with respect to such drug, there is lack of substantial evidence consisting of *adequate and well-controlled investigations, including clinical investigations* [emphasis added] by experts qualified by scientific training and experience to evaluate the effectiveness of the drug involved, on the basis of which it could fairly and responsibly be concluded by such experts that the drug will have the effect it purports or is represented to have under the conditions of use prescribed, recommended, or suggested in the proposed labeling.
>
> (ii) The following principles have been developed over a period of years and are recognized by the scientific community as the essentials of adequate and well-controlled clinical investigations. They provide the basis for the determination whether there is "substantial evidence" to support the claims of effectiveness for "new drugs" and antibiotic drugs.
>
> (a) The plan or protocol for the study and the report of the results of the effectiveness study must include the following:
>
> (1) A clear statement of the objectives of the study,
>
> (2) A method of selection of the subjects that (i) Provides adequate assurance that they are suitable for the purposes of the study, diagnostic criteria of the condition to be treated or diagnosed, confirmatory laboratory tests where appropriate, and, in the case of prophylactic agents, evidence of susceptibility and exposure to the condition against which prophylaxis is desired.
>
> (ii) Assigns the subject to test groups in such a way as to minimize bias.
>
> (iii) Assures comparability in test and control groups of pertinent variables, such as age, sex, severity, or duration of disease, and use of drugs other than the test drug.
>
> (3) Explains the methods of observation and recording of results, including the variables measured, quantitation, assessment of any subjects response, and steps taken to minimize bias on the part of the subject and observer.
>
> (4) Provides a comparison of the results of treatment or diagnosis with a control in such a fashion as to permit quantitative evaluation. The precise nature of the control must be stated and

an explanation given of the methods used to minimize bias on the part of the observers and the analysts of the data. Level and methods of "blinding," if used, are to be documented. Generally, four types of comparison are recognized:

(i) No treatment: Where objective measurements of effectiveness are available and placebo effect is negligible, comparison of the objective results in comparable groups of treated and untreated patients.

(ii) Placebo control: Comparison of the results of use of the new drug entity with an inactive preparation designed to resemble the test drug as far as possible.

(iii) Active treatment control: An effective regimen of therapy may be used as comparison, e.g., where the condition treated is such that no treatment or administration of a placebo would be contrary to the interest of the patient.

(iv) Historical control: In certain circumstances, such as those involving diseases with high and predictable mortality (acute leukemia of childhood), with signs and symptoms of predictable duration or severity (fever in certain infections), or in case of prophylaxis, where morbidity is predictable, the results of use of a new drug entity may be compared quantitatively with prior experience historically derived from the adequately documented natural history of the disease or condition in comparable patients or populations with no treatment or with a regimen (therapeutic, diagnostic, prophylactic) the effectiveness of which is established.

A summary of the methods of analysis and an evaluation of data derived from the study including an appropriate statistical method.

In practice, these regulations are usually interpreted to require a minimum of two adequate, well-controlled studies, preferably RCTs, for FDA to approve a drug for a particular indication. In October 1982, FDA published proposed revisions to the regulations (FR 47(202): 46622-46666) to further clarify the definition of "adequate and well-controlled investigations."

The drug approval process is without doubt expensive and time-consuming, facts that have not gone unnoticed by companies that develop and market drugs. The now infamous "drug lag," the long period that elapses between developing a drug and making it available to the public, has been blamed on lengthy testing. Arguments to extend the life of drug patents often point out that testing time so shortens the life of a drug sold under patent protection that companies are hard pressed to recoup their investment costs and make a profit before other drug companies market a "me-too" drug. *Patent-Term Extension and the Pharmaceutical Industry* (231) reviews the evidence and discusses the controversy on patent life.

The 1962 amendments require not only that new drugs meet safety and efficacy standards, but that all drugs approved between 1938 and 1962 be reevaluated by these criteria. The Drug Efficacy Study (DES) was set up to review the approximately 3,500 drug products still on the market of the approximately 7,000 that had been approved between 1938 and 1962. The National Research Council (NRC) of the National Academy of Sciences, carried out the DES between 1966 and 1969. The DES has been criticized for relying on "clinical experience," the very method of determining drug efficacy that the 1962 amendments sought to abolish (219). The DES found nearly 1,000 drugs to be ineffective, and most of the rest effective, at least for one indication. About 200 of the original 3,500 drugs remain to be finally evaluated, pending the completion of additional studies. FDA will assess these drug products as in new drug evaluations rather than as in NRC procedures.

While FDA closely regulates the introduction and labeling of new drugs, no one regulates the way drugs are used in practice. Although advertising must conform to labeling information, it is not uncommon for drugs to be used for many other indications than those specifically approved, and in dosages decided on by individual physicians. In practice, therefore, even though RCTs stand behind FDA's decisions to allow the introduction of new drugs, they may not stand behind decisions about how the drugs are used. To the extent that medical practice does not conform to RCT results, drugs may not be as safe and effective as they are presumed to be.

Overall, the drug approval process in this country has worked well. Drugs introduced since the 1962 amendments have not produced any disasters, and are probably effective. Reliance on RCTs for evidence of safety and efficacy must be viewed as a positive step. Adjustments may be made to streamline the drug approval process, but the need for adequate and well-controlled studies is immutable.

REGULATION OF MEDICAL DEVICES

RCTs play a role in FDA's regulation of medical devices. The 1976 Medical Device Amendments to the Food, Drug, and Cosmetic Act substantially increased FDA control over the safety and efficacy of medical devices. Safety and efficacy requirements apply to one of the three classes of devices named in the amendments: Class III devices, defined as those that are life-sustaining, life-supporting, implanted, or that present a potential unreasonable risk of illness or injury, and for which general controls or performance standards may not provide reasonable assurance of the device's safety and efficacy (234). These devices require premarket approval with information requirements similar to, but not as extensive as, those for approval of new drugs.

VETERANS ADMINISTRATION POLICY

The Veterans Administration (VA) Cooperative Studies Program (CSP) has not been geared to produce results specifically for VA policy, though its studies are selected for their relevance to the health of the veteran population. Hospitals and physicians in the VA system have the same freedom to decide on patient care as do hospitals and physicians in the private sector. Thus, VA distributes the results of CSP trials and trials carried out by other groups to their hospitals, but does not dictate that changes in treatment must occur as a consequence.

VA did base its decision to set up hypertension clinics on results which emerged from clinical trials. That decision was based on the pioneer studies of Edward Freis, a VA researcher, that showed the value of drug treatment of essential hypertension in preventing death from cardiovascular disease.

RCTs AND COST-EFFECTIVENESS ANALYSIS

"Decisions in the health care field are too often made on the basis of one option being more beneficial than another—irrespective of cost—or being cheaper and disregarding relative benefits; doctors were more prone to the first error, accountants to the second" (64). Greater use of some form of cost-effectiveness analysis (CEA) for making allocation decisions that affect the "medical commons" (110) should be a step forward for health policy. (For a complete discussion of CEA methods and uses, see ref. 229.)

The extent to which policymaking can rely on CEA depends in large part on the information available for the analysis. RCTs provide the soundest basis for the effectiveness side of the equation. Drummond and Mooney (64) mention several CEAs that relied on information from RCTs. One relied on an RCT of 2-day v. 7-day hospital stays after surgery for inguinal hernia or varicose veins, which showed no difference in patient outcome. CEA results showed the shorter stay to be more cost effective, though the saving was not as great as expected. Researchers have conducted other RCTs to study lengths of hospital stays, ambulatory compared to inpatient surgery, "cimetidine in the treatment of duodenal ulcer, the use of nurse practitioners in primary care, combinations of transplantation and dialysis in the treatment of chronic renal failure, and different methods of screening school children for asymptomatic bacteriuria" (64).

Recognizing the importance of the cost side of the equation, VA has begun to collect cost data in RCTs. Two VA CSP trials now in early stages of development are collecting data for CEA: one is a study of percutaneous transluminal angioplasty (of the femoral artery), the other of total parenteral nutrition in malnourished surgical patients. These studies will gather detailed information about all costs incurred in the treatments, including all visits to physicians within or outside the VA system. CEA features will also be encouraged in other appropriate new VA studies.

RCTs AND MEDICARE COVERAGE

The Medicare program came into being with the 1965 Social Security Act. Medicare, a nationwide, federally administered and funded health insurance program, provides benefits for people over age 65, for certain disabled individuals, and for those in certain other special categories. Because it is the largest health insurance program in the country, Medicare can influence the introduction and diffusion of health technologies through decisions about the benefits the plan will cover (229). In 1980, the Federal Government spent nearly $37 billion for Medicare, out of the total of $247 billion spent on health care in the country. RCTs already have had a small role in decisionmaking about what Medicare will pay for, and they may be much more widely used in the future. Ruby (194) states: "The rapid development of new and sophisticated technologies and the lack of specificity concerning benefits in most insurance plans, including Medicare, has led to the need for coverage determinations on a technology-by-technology basis." It is in such determinations that RCTs may be most useful.

The Health Care Financing Administration (HCFA) administers the Medicare program and is responsible for decisions about what medical services will be paid for, in keeping with the program mandate. The guiding principle in the law behind coverage decisions is that only those services that are "reasonable and necessary" will be reimbursed. No regulations define or delineate the bounds of "reasonable and necessary." In most cases, the fact that practices are widely used and accepted by the medical profession has been sufficient to ensure Medicare coverage. It would be impractical for the program to exclude from coverage all practices unsupported by RCTs. However, questions regularly arise about whether Medicare should cover a particular practice and some ground rules for making those decisions are necessary.

HCFA makes the final decisions about Medicare coverage, but relies on the Public Health Service to assess the medical and scientific aspects of health care practice, at HCFA's request. At present, the office that provides this service is the Office of Health Technology Assessment (OHTA) in the National Center for Health Services Research (Department of Health and Human Services), succeeding the short-lived (1978-81) National Center for Health Care Technology (NCHCT).

Most of these requests concern new technologies and new applications of existing technologies, though OHTA also looks at existing technologies suspected of being outmoded or of lacking effectiveness. As examples of the type of questions posed, OHTA has recently completed three assessments of apheresis for three different conditions, and is in the process of assessing that technology's use for three other conditions.

HCFA and most other third-party payers accept FDA's approval of a drug as the basis for coverage. Nearly all drugs marketed today have been through FDA's approval process, which is the most rigorous scrutiny of any medical technology in this country. (See section on FDA's approval of drugs.)

OHTA has drafted "Guidelines for the Evaluation of the Safety and Clinical Effectiveness of Medical Technologies" (237), which operationally addresses the "reasonable and necessary" criteria of the law. The guidelines state that three types of evidence are acceptable in deciding whether a technology meets these criteria: clinical trials, other well-designed clinical studies, and the medical opinion of qualified clinicians. Of the three, "most weight is given to controlled clinical trials or other well-designed clinical studies." Unfortunately, the results of RCTs have rarely been available for decisionmaking on the issues HCFA must resolve. On the 1982 list of 24 full-scale assessments for HCFA (table 5), RCT results were available only for two: the assessment of gastric freezing for peptic ulcer, which was done for historical interest and did not affect medical practice under Medicare (see box D in ch. 4) and the assessment of home blood glucose monitors (HBGM). The RCT of HBGM studied a total of 13 pregnant diabetics, 7 assigned to HBGM and 6 to urine glucose monitoring, with a control group of 8 nondiabetic pregnant women. The study found that HBGM was not essential for

Table 5.—Office of Health Technology Assessment Report Series 1982

Number	Assessments of Medical Technologies for the Health Care Financing Administration
1	Electrotherapy for Treatment of Facial Nerve Paralysis (Bell's Palsy)
2	Hyperbaric Oxygen Therapy for Treatment of Organic Brain Syndrome (Senility)
3	Hyperbaric Oxygen Therapy for Treatment of Multiple Sclerosis
4	Gastric Freezing for Peptic Ulcer Disease
5	Bolen's Test for Cancer
6	Bendien's Test for Cancer and Tuberculosis
7	Rehfuss Test for Gastric Acidity
8	Rheumatoid Vasculitis Therapeutic Aphersis
9	Home Blood Glucose Monitors
10	Ambulatory Blood Pressure Monitoring in Hypertensives (Semiautomatic)
11	Apheresis for Multiple Sclerosis
12	Hyperbaric Oxygen Therapy for Treatment of Arthritic Diseases
13	Plasmapheresis and Plasma Exchange for Treatment of Thrombotic Thrombocytopenia Purpura
14	Obesity and Protein Supplemented Fasting
15	Serum Seromucoid Assay
16	Percutaneous Transluminal Coronary Angioplasty for Treatment of Stenotic Lesions of a Single Coronary Artery
17	Melodic Intonation Therapy
18	Photodensitometry
19	Bone Biopsy for Mineral Analysis or Bone Histology
20	Photon Absorptiometric Procedure for Bone Mineral Analysis
21	Hyperbaric Oxygen for Treatment of Soft Tissue Radionecrosis and Osteoradionecrosis
22	Hyperbaric Oxygen for Treatment of Chronic Refractory Osteomyelitis
23	Carbon Dioxide Laser Surgery
24	Percutaneous Transluminal Angioplasty for Treatment of Stenotic Lesions of the Renal Arteries

SOURCE: Office of Health Technology Assessment, 1982.

good control of blood glucose in all pregnant diabetics. The remaining evidence on HBGM came from uncontrolled studies. The RCT did not play a major role in the study's conclusions.

In some cases of assessing practices, RCTs have played a dramatic role, but these are exceptions. An ongoing National Eye Institute trial of photocoagulation for macular degeneration concluded halfway through the trial that the procedure was effective. On the strength of the RCT, OHTA reversed its previous assessment that evidence of the procedure's effectiveness was lacking. HCFA now covers the procedure under Medicare.

OHTA keeps an eye on ongoing trials to act quickly when decisive information becomes available. One current trial that could affect Medicare policy is one of apheresis for systemic lupus erythematosus.

Overall, RCTs have not been used in testing many practices of concern to HCFA. According to Seymour Perry, former head of NCHCT, "the NIH [National Institutes of Health] infrequently supports clinical trials designed to answer the kinds of specific questions that are embodied in technology assessments" (177). RCTs that are carried out may fail to answer questions of interest to HCFA. First, RCTs do not always compare competing technologies but often only assess the safety and effectiveness of new individual technologies. In making policy, however, it is often better to compare competing technologies directly. Trying to compare separately conducted RCTs of two or more competing technologies is exceedingly difficult. Differences between the patient populations and the study designs may make the comparison of studies all but impossible.

Second, the Medicare population, mainly the elderly, is not always represented in RCT patient populations. Medical interventions often have different effects on different age groups, and the results of an RCT including mainly those under 65 may not be directly applicable to the Medicare population.

Of interest to policymakers in general is the effectiveness of medical technologies under conditions of normal use. Treatments are usually more strictly controlled in RCTs than is possible in usual practice. This is a third drawback to applying RCT results directly to policy decisions.

A fourth problem is lack of timeliness. Results of RCTs often are long in coming, and may lag behind changes in practices, especially the introduction of new procedures. HCFA often cannot wait for RCT results. When results do become available, HCFA may change its policies accordingly. This is relatively easy if the change is from noncoverage to coverage. In the case where an RCT provides evidence counter to the use of a technology for which coverage has already been granted by HCFA, a reversal is more difficult. Greater evidence would be needed to refuse payment for a technology once permission to use that technology has been given than the evidence that could accomplish the same in initial decisionmaking (254).

THE POTENTIAL IMPACT OF RCTs ON THIRD-PARTY PAYERS

From the early years of Medicare until quite recently, new procedures endorsed by the medical community were reimbursed with little questioning, and with no requirement for sure evidence of efficacy (7). One can assume that a certain proportion of medical practices are in fact not effective. Evidence from RCTs that demonstrated a practice lacked effectiveness could theoretically put an end to the practice, perhaps cutting the costs to Medicare and other third-party payers, without eroding the quality of medical care. While RCT results are not unassailable by proponents or opponents of particular practices, they provide a much sounder basis for decisions than do other kinds of evidence.

The impact of RCTs on coverage decisions by Medicare and other third-party payers will depend on the result of the RCT and the way in which the information is used. Studies providing convincing evidence that a technology is not effective should be the easiest to incorporate into coverage decisions. Denying coverage for an ineffective intervention will both save money and save people from undergoing treatments that will not help them. The potential for cost-savings is substantial. An analysis of the savings from four decisions for noncoverage made by HCFA indicates that the Medicare program was saved between $88 million and $959 million over a 10-year period, presumably with no loss of clinical benefit (7).

Not all RCTs provide negative evidence. Some things work; they are safe and effective. Effectiveness is not the only criterion for coverage by any third-party payer, however. It may not be "reasonable and necessary" for Medicare to provide artificial hearts to all who might qualify for them, for instance. Other factors, notably cost, may render an effective technology unreasonable. Private third-party payers have greater freedom to extend or deny coverage than does Medicare. Private organizations may be more responsive to market supply and demand in what they offer. They may trade lower premiums for more limited coverage. The Medicare program does not have that option. The use made of positive results from RCTs will probably vary more than will the use of negative results. In either case, however, decisions made in the light of results from well-designed, well-conducted RCTs should be more rational, less subject to chance than decisions made without such results.

Bunker and Fowles (27) have proposed one mechanism for generating clinical information that would be useful to a variety of decisionmakers, including third-party health insurers. Their model is a centralized Institute for Health Care Evaluation (IHCE) (see box C) which would be supported by health insurers, but would work independently in funding research, including RCTs. The aim of IHCE would be to provide decisionmakers with information on which to base coverage decisions.

Box C.—Institute for Health Care Evaluation

Bunker and Fowles have proposed a model for an Institute for Health Care Evaluation (IHCE) (234). The goal of IHCE would be to "generate cost-effectiveness data with a strong emphasis on the measurement of outcomes of therapeutic intervention." A major IHCE activity would be to generate new information, through the support of clinical trials, when appropriate. Proposed membership in IHCE includes private and public third-party payers, health maintenance organizations, professional associations, and health care consumers.

An advantage of an independent institute is that it would insulate technology assessments from undue influence by interested payers. Because third-party payers do have a stake in the outcome of assessments, more direct participation in funding RCTs could raise questions of conflicts of interest.

Financial support from insurer members could be voluntary, or perhaps, mandated as a tax through new legislation. Each avenue presents both advantages and disadvantages.

Under the taxation approach all health plans (for-profit and nonprofit) would be required to contribute according to some per-capita or other formula. This would eliminate the problem of "free-riders" (i.e., competing programs which gain access to information without paying for the costs of its generation).

A voluntary mechanism, while a less secure approach to funding, might be more palatable to insurers, particularly in getting the Institute established. A system of voluntary contributions might be more vulnerable to pressures from members concerning the activities of the Institute, however.

4.

Factors Affecting the Impact of RCTs on Medical Practice

4.

Factors Affecting the Impact of RCTs on Medical Practice

The decision to conduct a randomized clinical trial (RCT) creates a potential impact on medical practice. The act of participating in a trial may have a limited impact, on the practice of at least those physicians directly involved. Once an RCT is complete, both its own characteristics and those of the technology it is used to evaluate determine the trial's impact. This chapter first outlines the objections and alternatives to RCTs that may bear on the decision to carry them out. The latter part of the chapter describes those characteristics of RCTs that appear to most influence their impact, which include:

1. the timing of the trial with regard to the technology's degree of development and diffusion;
2. the constituency supporting the technology prior to the trial;
3. the quality of the trial, both in statistical and other design features;
4. the fact of whether the trial is conducted through one or more centers;
5. the form of disseminating trial results; and
6. other important characteristics.

OBJECTIONS TO RCTs

Objections are rarely if ever raised to the principles of controlled experimentation on which RCTs are based. RCTs themselves, however, are not universally accepted. Two objections are commonly raised against them:

1. that they are too difficult to conduct; and
2. that they may violate the ethical principles that apply to all experimental research involving human beings.

Practical Problems in Conducting RCTs

Objections to RCTs because of their practical problems focus on the use of resources. RCTs are expensive compared with other study designs, can require long periods of followup, and can be administratively complex. If other study designs could answer the questions asked as RCTs can, these objections would be compelling. This is not the case, however, as a later part of this chapter explains ("Alternatives to RCTs").

With regard to cost, it is easier to put a price tag on an RCT than on the expense of not doing one. The widespread adoption and use of ineffective technologies can waste scarce resources. For instance, before a great deal of diffusion, RCTs checked the use of hyperbaric oxygen treatment for cognitive deficits in the elderly, a practice that could have become widespread (see box F in ch. 5). The balance sheet for RCTs might look different if their "credits" could be shown as easily as their "debits." This is not to claim that every RCT saves money in the long run.

RCTs, especially multicenter RCTs, can be complex administratively. Like all other good research, they require careful planning, execution, and data handling and analysis. These do not appear to be valid reasons for not undertaking RCTs. To some extent, the pratical problems have been lessened by the widespread availability of computers for data handling.

Ethical Issues in Conducting RCTs

The most frequent objections to RCTs are on ethical grounds. These objections center on the rights of patients to get the best treatment available and the responsibility of physicians to provide it. Clearly, certain kinds of experimentation on human beings are not acceptable. When the evidence is overwhelming that a newly developed therapy is efficacious—penicillin for pneumococcal pneumonia, for instance—it would be unethical to withhold this therapy from a control group, although RCTs might be appropriate to determine its optimal regimen. The choice between competing technologies or the superiority of an innovation is not always clear. Ethical issues are most difficult in the middle ground where uncertainty is greatest.

The decision to fund an RCT, or any human research, involves at least an implicit decision that the trial is ethical and that it addresses an important question about which uncertainty exists. After this point, the mechanism protecting the individual's rights are procedures of "informed consent." While informed consent may appear a simple idea, universally acceptable methods of seeking and obtaining informed consent still elude us, though progress has been made.

International bodies have developed ethical codes addressing the particular problems of research. Such codes include the Nuremburg Code and the World Medical Association's Declaration of Helsinki (13). In this country, the Department of Health and Human Services (the Department of Health, Education, and Welfare prior to 1979) has conducted a number of studies on human research that offer guidance on ethical issues (132).

A number of other measures have been proposed to minimize the subtle coercion of patients in obtaining their consent to participate in trials. For example, the World Medical Association suggests that a physician who is not part of the investigation discuss informed consent with the patient. The Department of Health, Education, and Welfare's National Commission recommends giving patients adequate time to decide whether to participate and reducing other potentially coercive environmental conditions (132).

The Cancer Research Campaign Working Party in Breast Conservation recently recommended the following points to improve methods of seeking in formed consent in breast cancer trials (33):

> (1) Eligible patients should be given the option to take time to consider giving their consent, perhaps along the lines described by Simpson at the Wellington Hospital in Australia [77]. Here the patient is fully informed about the trial by her physician or surgeon but an informal consent in principle only is obtained. At a later date the procedures are again explained and only then is formal consent obtained by asking her to sign a consent form.
>
> (2) The consent form should be fairly non-specific but it must be backed up by as much verbal explanation as possible. Signature to such a form in the presence of a witness might have legal validity if it included the phrase "the effect and nature of such treatment have been explained to me," but only if it could be proved that the explanation had been given [68].
>
> (3) Ideally, a trained nurse counselor or other suitably qualified person should help to obtain informed consent, and the patient should be made aware that she may resume this continuing dialogue at any time.
>
> (4) Ethical committees should view the issue of informed consent as a top priority, bearing in mind its various applications—in the ordinary clinical situation, in therapeutic trials, and in experimental research. They should reconsider the type of guidelines to propose to doctors, with reference to the Declaration of Helsinki and other national and international codes and regulations; they should consider practical ways of improving consent procedures in their hospitals; and they should monitor these procedures, perhaps by requesting reports at stated intervals.
>
> (5) Those doctors who treat patients with cancer but do not participate in randomized clinical trials should realize that they too have an obligation to discuss alternative forms of treatment with their patients. In our view the fact that they are not formally randomizing their patients does not reduce their obligation in this respect.

While RCTs in this country today require the "informed consent" of the participants, the procedures used to obtain consent vary considerably. Critics point out that the rights of certain classes of patients, e.g., children, the aged, the mentally retarded, and prisoners, are easily violated. The steps taken to protect patients' rights are generally reviewed by at least the funding organization and any institutional review board with jurisdiction over the investigators. While patients' rights are a major concern, mechanisms have been established to protect those rights.

A unique issue arises in seeking patients' informed consent to participate in an RCT. Seeking consent for a particular procedure is more easily accomplished than seeking consent to be randomized and to undergo the uncertainties of such assignment. The role of and need for randomization must be communicated, as well as the risks and benefits of all possible treatment assignments.

The difficulties of seeking informed consent for RCTs may be daunting, but they are not reasons to abandon RCTs. If the same standards of informed consent were applied to experiments with control treatments, the difficulties might appear less (33):

> This argument may be taken to its logical conclusion: that clinicians treating patients outside any protocol in any area of controversy also have the obligation to inform their patients of the alternative treatments that are being offered in different parts of the country at the same time.

A common contention is that control groups are deprived of the benefit of therapy by participating in an experiment. It is a common misconception that control groups are administered only a placebo or no treatment at all. In any case where a standard or accepted technology is challenged by a new technology, the ethical comparison is usually between the standard and the new. In some cases, it may be ethical to use a placebo even if some treatment is available, for instance, in a trial of a new headache remedy, a placebo might be used instead of aspirin.

A frequent objection to randomizing is that some patients will be denied access to the innovative intervention. On the other side of the coin, objections may be raised because participants are subjected to new technologies with unknown risks (21). A general conviction that research subjects are exploited or manipulated regardless of the benefits they might receive contributes to the ethical objections (21).

The responsibility of a physician is to give patients optimal treatment. Ethical arguments against randomizing state that physicians should act on the best information available and choose the intervention they believe is superior. When uncertainties about new or existing interventions allow no clear distinction, "a physician makes the intellectually honest admission that best therapy is not known, and than an ethical course of action is to undertake a randomized clinical trial to find out" (32). In fact, the ethical failure of relying on uncontrolled experiments is that lack of effectiveness and side effects are recognized much later than they would be if tested in RCTs (33).

Ethical issues may confound attempts to evaluate practices that are questionable but so entrenched in medical practice to make an RCT all but impossible. Hiatt (110) cites as examples coronary care units (CCUs) in hospitals in this country, and cytologic screening for cervical cancer. Treatment in a CCU indisputably adds greatly to the cost of care but is of unknown value in lowering mortality from myocardial infarction. The development and subsequent widespread use of cytological screening for cervical cancer (the Papanicolau or Pap test) followed a decline in the incidence of that cancer. The value of this screening and the optimal interval for its use are unknown. Both these interventions use a great deal of health care resources: the first mainly because each episode of its use is costly, the second because it is applied to almost half the adult population, and up to 40 or more times during the course of each woman's life. In the case of CCUs, two RCTs in Great Britain found no advantage of CCUs over home care. Nonetheless, RCTs in this country would be extremely difficult to do, and if results were contrary to current practice, they would probably be received unfavorably.

Ethical concerns do not disappear once a trial starts. As data are continually gathered and endpoints recorded, answers about safety and efficacy may emerge more quickly than anticipated. In the case of detecting unsuspected adverse effects, as occurred in the Coronary Drug Program (ch. 5, "RCTs in Cardiovascular Disease") and the University Group Diabetes Project, a decision must be made about when to discontinue treatment. In such cases, however, there are no rules to rely on. Differences of opinion arise about questions of safety as well as of efficacy. Some investigators will be convinced earlier than others that one therapy is better than another. Decisions to stop large-scale trials are generally made by an oversight committee of some sort, and are reached by consensus. Klimt (124) discusses the major issues involved in terminating a long-term trial.

Whether enrolled in a clinical trial or not, a patient deserves the best possible treatment from his or her physician. Particularly in long-term chronic disease trials, patients' conditions may change during the course of the trial so that different treatments are indicated. In whatever way a trial is organized, a physician retains and must exercise the responsibility to withdraw the patient at any time, or to offer the competing treatment or a different one, whenever any such change is in the patient's interest. In recent trials of coronary artery bypass surgery that assigned individuals to surgical or medical treatment, a large number of those assigned to medical treatment have subsequently undergone surgery for intractable angina. These necessary changes of treatment have changed the research question from "Which is more effective, medical or surgical treatment?"to "Which is more effective, immediate surgical treatment or immediate medical treatment, followed by surgery only in those patients for whom medical treatment is insufficient?" The second question conforms more closely to actual practice than the original one.

Another ethical concern is how long researchers and funding agencies should follow those patients who participate in clinical trials (255). Perhaps a lifetime followup is desirable for some classes of participants. The potential long-term effects of some chemotherapeutic agents are worrisome, especially those of anticancer drugs. At present, funding agencies do not routinely provide for long-term followup.

ALTERNATIVES TO RCTs

The money and time that RCTs require have led to a continued search for alternative means to determine the safety and efficacy of medical technologies. It is generally argued that any acceptable method must compare a group that undergoes the new treatment (or other intervention except in the rare case that the experimental treatment is an obvious major breakthrough) with a group that does not. The arguments center on the ways in which these groups are assembled. The major rival to RCTs has been the type of study that uses "external controls," most frequently "historical controls." External controls are those drawn from populations that may differ, in unknown ways, from the study population. Historical control trials (HCTs) compare a group of patients treated by the new intervention with a group treated sometime in the past in another way. Another type of external controls are patients treated during the same time period at the same or different institutions from the experimental group, but who are not assigned to treatment according to the experimental plan ("concurrent controls").

The data on historical controls ranges from dim personal remembrances to that gathered carefully and in detail by investigators (24). Historical controls may have been treated at the same or a different institution as the experimental group. They are generally chosen from the literature, from the immediately preceding trial in a sequence of trials, or matched from a previous study (88). Successful matching assumes knowledge of important prognostic factors, which is often not a valid assumption.

The attractions of historical controls are several. HCTs sidestep the question of whether it is ethical to randomize patients. Studies with historical controls require the active cooperation of fewer participants since data need be newly collected only for the experimental patients. Requiring fewer participants makes studies proportionately cheaper. Recruitment into the study is improved to the extent that patients need not consent to randomization and are sure of the treatment they will receive beforehand.

Gehan and Freireich (88) argue that clinical trials in cancer research should sometimes use a selected rather than a randomized control group. They cite the following kinds of cases: 1) when the study attempts to determine the absolute rather than the relative effectiveness of the treatment, 2) when large differences in response rate between treatment groups are based on preliminary trials, and 3) when a therapy can be compared to a standard therapy evaluated in a recent trial. Addressing at least the second kind of case, Chalmers, Block, and Lee (45) argue for randomized controls on the grounds that most drugs tried to date in cancer therapy have been relatively ineffective.

In the past, data on external controls have usually been gathered from patient records by abstracting the relevant information. Because the primary purpose of such records is for patient care rather than research, the requisite information is often not recorded. Data banks are a relatively new development that may improve the quality of external controls, but this is yet unproven. Medical data banks are usually created by establishing a common vocabulary to describe clinical histories, and then observations on patients are entered as events occur (234). The uniform information available about patients can be used to improve the comparability of an experimental group and a group of controls (who are chosen from a data bank). Nevertheless, data banks do not solve the problem of treatment changes over time that may render groups incomparable, particularly because not all medically significant variables can be identified. While data banks may be useful in discovering some important prognostic factors, they are not good enough to compare treatments (99). In this regard, Byar observes (31):

> The great danger seems to me to be that data banks will be seen as a replacement for randomized trials, whereas in fact the most useful data which could be stored in data banks would be those obtained from randomized studies.

When a technology is so widespread or well established that use of untreated controls would be questionable, investigators must then rely on historical data. When random assignment to groups is possible, however, the available evidence suggests it is superior. Wortman and Saxe (252) compare the validity of RCTs with that of HCTs (and other epidemiologic study designs). The major advantage of RCTs is their internal validity; i.e., high probability that the effects they reveal result from using the technology and not from some other factor. HCTs, in contrast, often lack internal validity. Whether identifiable or not, changes over time in medical practice or the patient population are often equally likely explanations of effects detected in HCTs. This is illustrated by changes in the treatment of osteogenic sarcoma. The history of this treatment points to the hazards of comparing aggregate survival rates from time periods before and after a procedure is introduced (252):

> Following the development of this treatment in the early 1970's, researchers began to experiment with ways to treat patients with the drugs before their cancer metastasized. Historical controls drawn from patients' records dating from the 1960's were used in this research, and the results provocative. Nearly half the patients treated lived 2 years without a recurrence of the disease, compared to only 20 percent of the patients in 1960.
>
> Unfortunately, the change in therapy from 1960 to 1970 was also accompanied by other changes in diagnosis, treatment, and patients. The use of the computed axial tomography (CAT) scanner in the 1970's provided a much more sensitive test for detecting patients who did not have metastases. At the same time, surgeons began removing metastases in the lungs. At the Mayo Clinic, where both of these techniques were employed without chemotherapy, the survival rates equaled those of patients treated with the drugs.
>
> In addition, the patient mix probably changed over time so that those with the worst prognosis no longer constituted the majority of those treated. These criticisms of the research design and findings of a small controlled trial have convinced the National Cancer Institute to support a multicenter RCT to assess the efficacy of adjuvant chemotherapy for osteogenic sarcoma.

Sacks, Chalmers, and Smith (197) compared the outcomes of RCTs for six therapies that each had been tested by at least two RCTs and two HCTs. In every case, HCTs indicated these therapies were more beneficial than did RCTs, the difference lying mainly in the outcomes of the control groups. In HCTs, control groups fared considerably worse than controls in RCTs, while the treatment groups fared about the same. To provide a better comparison, the results of some HCTs were adjusted to account for differences in prognostic factors between HCT and RCT groups. Sacks and colleagues found that this had little effect on the analysis and concludes that little can be done to improve the accuracy of HCTs. The problem of using historical control is not the existence of bias per se, but the impossibility of detecting, measuring, or removing it.

HCTs are more likely to favor a new treatment because of the nature of historical controls. RCTs are more likely to find no difference between treatments even if a difference exists. Although other factors may contribute to not detecting an effect when it actually exists, the main culprit is an inadequate sample size, and not an inherent weakness of RCTs. The problem could partly be solved by greater emphasis on power considerations in experimental design, with planning for sample sizes large enough to ensure finding any important difference in treatment groups.

Sacks and colleagues (197) suggest in addition that the "nearly automatic" use of a p value of 0.05 as a measure of statistical significance may not always be appropriate. Such an association means that the prespecified result is expected to occur by chance alone 5 times out of 100, given the sample size of the trial. They suggest that positive results of RCTs might be accepted as true positives even assuming a greater possibility that the results may be due to chance. On the other hand, given the bias in favor of new interventions in HCTs, a more stringent significance level might be required of them for the same level of proof.

Wortman and Yeaton (253) synthesized the results of studies of coronary artery bypass graft (CABG) surgery. They looked at both RCTs and nonrandomized studies with concurrent controls reported between 1970 and 1981. They conclude that both kinds of trials favor surgical treatment, but that nonrandomized studies tend to overestimate its benefit. They combined data on survival and mortality from 9 RCTs and 16 nonrandomized studies by means of two different synthesis techniques. In both cases they found that the average benefit to the surgical patients as computed from nonrandomized studies is four to eight times greater than that computed from RCTs.

Studies to date comparing RCTs and other types of studies indicate that RCTs are and should be the favored method for evaluating major clinical recommendations and should be abandoned only when special conditions preclude them.

CHARACTERISTICS OF RCTs THAT AFFECT THEIR IMPACT

Timing of RCTs

At what point in the life of a medical intervention should it be tested in an RCT? The law and regulations answer this question for new prescription drugs and vaccines, requiring RCTs of nearly all. The safety and efficacy of pharmaceuticals must be demonstrated before they can be widely used. To other kinds of interventions, e.g., surgical and radiological ones, no such law applies. RCTs have typically been initiated when a critical amount of skepticism has developed about the effectiveness of an intervention. By then it may have attained widespread popularity, with its attendant consequences—e.g., major investments in learning skills, such as surgical techniques, or in equipment. Many people have been subject to an intervention of unknown efficacy, including ineffective ones, such as gastric freezing for duodenal ulcer (see box D) and some that are actually harmful. These problems may be confounded by the usual delay inherent in changing even a bad technology, and the increased grounds for malpractice suits for an abrupt public admission of error.

One approach to the timing of trials is to "randomize the first patient." Chalmers is one of the main proponents of randomizing patients to treatments with the first use of a new intervention. He cites several times this has occurred, including trials of prophylactic use of portacaval shunt surgery (a procedure to allow blood flow to bypass the liver) for portal hypertension (abnormally high blood pressure in the veins of the liver, a frequent complication of liver cirrhosis) and colon bypass for chronic encephalopathy (a degenerative disease of the brain) in patients with cirrhosis (41). Randomizing from the very first is possible in some cases, but there are convincing arguments to delay the start of RCTs (though not to delay establishing formal systems to collect data).

BEST COPY AVAILABLE

Box D.—Gastric Freezing

The rise and fall (1962 to 1969) of "gastric freezing" in treating duodenal ulcer is a classic story. The procedure consists of a patient swallowing an uninflated balloon to which tubes are attached. Once in the stomach, the balloon is filled with a coolant, maintained at -10° C for about an hour, after which the balloon is deflated and removed. Claimed by its originator, Owen Wangensteen, a leading academic surgeon, to decrease gastric secretions, to relieve pain, and to be safe, simple, and relatively inexpensive (245), gastric freezing quickly gained popularity. The only rival treatment to gastric freezing was palliative medical treatment with antacids, sedatives, and changes in living habits, or in severe cases, surgery with a mortality rate of 5 to 10 percent (160).

Despite enthusiastic adoption of gastric freezing, enough doubts about it remained to spur the planning of a multicenter RCT in 1963. When the results appeared in 1969 showing no difference in outcome between the group that had received gastric freezing and the group given a sham procedure, 2,500 gastric freezing machines were in use. According to Miao, the convincing results of the trial led to rapid abandonment of the procedure (160).

In a somewhat different interpretation of the events, Fineberg suggests that even before publication of these results, gastric freezing was on its way out. The negative result of the RCT, he claims was "of little practical consequence, as if a marble tombstone were erected over the grave of a patient already several years deceased (71)."

Arguments in favor of early RCTs are supported by the use of untested interventions later proved either ineffective (e.g., bed rest for hepatitis, the Sippy milk diet for gastric ulcer [40]) or harmful (e.g., prophylactic portacaval shunt surgery for portal hypertension, which was both ineffective and caused a type of brain damage in some patients [27]).

Doubts have been raised about the efficacy and safety of some technologies, yet years pass before they are tested in RCTs. Radical mastectomy was introduced around the turn of this century. In 1948, the simple mastectomy was proposed as an alternative. RCTs, which demonstrated the equality of the two procedures in patient survival rates, waited until 1969 and 1973. RCTs of bed rest for hepatitis, a bland diet for peptic ulcer, and diethylstilbestrol to prevent spontaneous abortion were delayed for similar periods of time (40).

Three facts argue against very early RCTs of surgical procedures. First, as surgeons' skills in performing a procedure improve, the results of performing it may improve, as measured in mortality or morbidity rates. Second, as experience accumulates, improvements to the procedure itself will be made, not only by clinicians involved in trials but by other practitioners. If the procedure evolves to a somewhat different and improved form, the ethical and methodological question arises whether a trial in progress should continue. The Veterans Administration's (VA) RCT of CABG surgery was a well-designed trial, but had minimal impact, in part because changes in techniques made the results irrelevant to practice by the time the trial had ended (20). In this trial, the procedure initially used, the Vineberg implant, was replaced with the newer CABG surgery. Data analysis was further complicated by a higher rate of operative mortality in the earlier CABG patients compared with the later ones. Third, when an innovation is better known, it may be applied to a changing set of patients. In particular, a promising but risky therapy may be applied to patients in earlier stages of disease, patients who may in fact benefit more from the procedures because they may have not yet begun to suffer some permanent late effects of the disease.

Bonchek (20) cites two well-designed RCTs in which problems arose because of the trials' delay in relation to the diffusion of the technology. The Coronary Artery Surgery Study began in 1974 after much experience with the procedure had accumulated. Excluded from the study were some high-risk patients of great interest (e.g., those with unstable angina). By the time the study began, their physicians presumably preferred them to have surgical treatment. Recruitment into the study was slower than expected, so the enrollment period was extended. Such delay in recruitment creates its own problems owing to evolutionary changes that take place, as was discussed above. A similar problem in recruiting patients occurred in a single-center study of unstable angina at the University of Oregon. Recruitment declined as physicians diverted their patients from the university hospital, not wanting them to be randomized.

Problems with the timing of trials are difficult, and there are advantages and disadvantages to carrying out trials at specific points in the diffusion process. In general, however, the arguments for earlier trials are stronger. The earlier RCTs occur, the sooner sound information is available for medical decisionmaking. The examples mentioned of "late" RCTs, and of no RCTs at all (for most current procedures) are more typical than those of RCTs conducted too early.

The Constituency Behind the Intervention

A strong interest group obviously supports the trials of new drugs. Those with a financial stake in these trials see that the results of positive ones are translated into practice as widely as possible. There is a general consensus that the results of positive drug trials are disseminated widely, and that physicians rapidly adopt new drugs. If there is any problem in adopting new drugs, it is their overuse. Although drug companies cannot label their products for indications other than those for which they have been given FDA approval, physicians are not bound by any law to prescribe according to RCT results.

When RCTs of already marketed drugs have negative results, the situation can be quite different. Beginning in 1961, the University Group Diabetes Program (UGDP) tested a popular hypoglycemic drug, tolbutamide, used in treating adult-onset diabetics to control their blood glucose. Early results of this trial indicated that the drug was unsafe (see box E), and the corresponding part of the trial was discontinued. This finding on tolbutamide set off a heated debate, which is now 13 years old and still alive.

Procedures also have their constituencies. The developers of new procedures and techniques have a professional stake in having them accepted and widely used. Financial interests may also be present when capital equipment is involved, e.g., imaging equipment or devices like heart valves, and joint implants, and when procedures are regarded as high reimbursement items by third-party payers. Positive results seem to have a greater impact in these cases than negative results. A potentially beneficial new procedure is welcomed by practitioners, particularly when the condition it treats is life-threatening and there is no alternative treatment. Rather than abandon a procedure for no treatment, even if an RCT shows little or no benefit, physicians may prefer to continue what they see as the only hope.

The Quality of RCTs

"Quality" in research cannot be precisely and categorically defined but criteria can be established to measure some of its features. Bailar (6) suggests two methods to judge quality: 1) evaluating the quality of the published research report, and 2) evaluating the quality of the work itself. Publications concerned with the quality of RCTs have taken both approaches. Regardless of whether better quality RCTs will have greater impact than those of poor quality, on general principle it is worthwhile to ensure that they are of the highest quality possible.

Most writers who focus on the quality of RCTs use the published literature as their source of data. Some have reviewed published RCTs to determine what features of the trials are reported, with the aim of judging the quality of the published reports. Others have taken data from these publications, i.e., the number of participants and other quantitative items, to judge the quality of the research. These two types of evaluations are discussed below.

The Quality of RCT Reports

Chalmers and colleagues propose a method to evaluate the quality of published RCTs, and a quality index based on this evaluation (47). They give heavy weight to the form of blinding, including blinding during randomization, that of physi-

Box E.— The University Group Diabetes Program

In 1961, the University Group Diabetes Program (UGDP) began an RCT "unique in the amount of rancor it has aroused and the length of time it has lasted" (142). The trial was sponsored by the National Institute of Arthritis, Digestive, and Metabolic Diseases, to settle longstanding questions about the treatment of "adult-onset" diabetes. The disease is characterized by the impaired ability to metabolize carbohydrates, stemming from the inefficient use of endogenously produced insulin. Traditionally, treatment consisted of controlling blood sugar (glucose) levels by injections of exogenous insulin, dietary management, or taking oral hypoglycemic drugs (agents that act to lower the level of glucose in the blood). The actual value of controlling blood sugar, however, was unknown. Two schools of thought were prevalent at the time: one holding that strict control was warranted, the other that the discomfort, inconvenience, and anxiety of strict control were not worth its benefits (142).

One aim of the UGDP RCT was to evaluate the control of blood glucose on the development of major complications of diabetes, particularly atherosclerotic heart disease, the most common cause of death among diabetics. The trial also set out to study the natural history of complications of the disease and to improve methods in clinical trials.

About 1,000 patients in 12 centers were instructed in dietary control and randomized to one of four treatments: 1) insulin in variable dosages to keep blood glucose at specified levels, 2) insulin in fixed dosages, 3) tolbutamide (an oral hypoglycemic agent widely used at that time), and 4) placebos in the same form and scheduling as tolbutamide. A fifth group, receiving a new oral hypoglycemic agent, was added after the study had begun.

The trial employed rigorous techniques of data collection and patient evaluation, relying whenever possible on objective measures of pathology and functional impairment. Many of these quality assurance and control measures had never before been employed in a large-scale trial. The followup was scheduled to last 10 years.

By the end of the eighth year, higher rate of cardiovascular mortality, one significantly higher than in any other group, had occurred in the group taking tolbutamide. The investigators discontinued its use and announced the results, touching off a controversy still unresolved. Their further conclusion, that insulin was no more effective than dietary control alone in preventing fatal vascular complications, added fuel to the fire.

A hue and cry arose from diabetologists, drug manufacturers, and publishers who carried advertisements for the drugs. The study was scrutinized and attacked on two major counts: 1) that treatment of the participants in the trial did not measure up to standards of clinical practice at the time; and 2) that a failure of randomization placed more high risk individuals in the tolbutamide group than in the others, rendering the results invalid.

In response, the National Institutes of Health (NIH) reviewed the trial and found it valid. The Biometric Society undertook a 2-year review of all the statistical aspects of the trial and came to the same conclusion. The Food and Drug Administration (FDA) conducted a 2-year audit, visiting the treatment sites and checking the data. They found no error (43). The data were finally reviewed by the courts during 10 years of legal action against the principal investigator. The UGDP trial is surely one of the few whose data have been found satisfactory by the Supreme Court of the United States.

The UGDP results were published in 1970. Not until 6 to 8 years later did sales of hypoglycemic agents begin to decline (43). In this case it may take the emergence of a new generation of physicians and patients for the practice to change entirely. One effect of the trial may be the policy decision of drug companies not to develop new hypoglycemic agents; none have attempted to seek approval for such agents since the controversy started.

Aside from its medical conclusions, the UGDP led to great debate about the value of RCTs in general, and revived the old issue of the relative value of inference and clinical judgment.

BEST COPY AVAILABLE

cians and the patients with regard to the therapy given, and that of physicians with regard to ongoing results. Analytic techniques, control of bias, description of patient population and treatments, and various aspects of quality control are counted as well. Adherence to the standards set down by these authors might raise the quality of RCTs, and might also facilitate comparing and synthesizing the results of small trials, particularly those with conflicting results.

DerSimonian and colleagues (62) studied the quality of reports of RCTs in 67 articles published in the *New England Journal of Medicine,* the *Journal of the American Medical Association,* the *British Medical Journal,* and the *Lancet* during specified time periods in 1979 and 1980. They chose 11 items of methodological importance and determined how often each was reported. A low score might indicate a poorly conducted study, a poorly reported one, or both. Information about statistical analyses, the names of statistical tests used, and the fact of random allocation to treatment were relatively well reported—at least 80 percent of the articles mentioned these items. Only 19 percent reported the method of randomization, 37 percent the eligibility criteria for admission to the trial; 57 percent whether patients were blinded, and 30 percent whether those assessing outcomes were blinded. The least frequently reported item was the statistical power of the trial to detect differences in outcomes, which was reported in only 12 percent of the articles. There were substantial differences among the journals, but they were as great as within-journal variation among articles. DerSimonian and colleagues conclude that journal editors could influence the quality of published trials by setting standards for reporting. The items of information they identify as important should be available to all authors, and could theoretically be reported 100 percent of the time.

Mosteller, Gilbert, and McPeek (165) came to similar conclusions in their reveiw of RCTs in cancer research. They looked at the frequency of reporting of five statistical and two procedural aspects of trials: randomization, statistical method, blinding, statistical power, sample size, patient survival rate, and informed consent. Each item was in 0 to 50 percent of the articles, with "24 percent as a reasonable overall single-number summary," of the frequency any item was reported. (The authors' recommendations based on this study are discussed in ch. 6.)

Haines (103) notes a number of deficiencies in reports of RCTs in neurosurgery, in addition to low statistical power. He found inadequate descriptions of blinding, of interventions tested, and of the eligibility criteria used. Haines did note a trend, though weak, toward improved quality over time, as determined by the scoring system of Chalmers and his colleagues (47). The participation of a biostatistician in the study, as evidenced by authorship or acknowledgment, was the most important correlate of whether a study was judged of good quality.

The Quality of RCT Research

Hemminki (107) cites 29 reviews on the quality of clinical trials published between 1950 and 1977. Hemminki's work was prompted by her previous review of clinical trials submitted to the drug licensing authorities of Sweden and Finland, which showed many trials to be both poorly reported and poorly done. Her conclusion echoes that of the authors of the original reviews, namely, that the majority of published trials were inadequately controlled or otherwise methodologically inadequate. Among the common deficiencies she cites, e.g., lack of statistical power, and lack of information about randomization and blinding techniques, Hemminki includes the unsatisfactory cojoining of information about adverse effects and beneficial effects. Adverse effects, which are generally rare, are usually analyzed separately from indications of effectiveness in comparing therapies. Hemminki suggests expressing both adverse and beneficial effects using the same scale, as in cost-effectiveness analyses. The most frequent criticism of many RCTs is that their sample sizes have been inadequate. Combined with other factors, small sample sizes lead to trials that have little power to detect moderate differences between groups. Statistical power and statistical significance in RCTs are discussed after reviewing other issues of quality in their design, execution, and analysis.

The use of appropriate statistical tests, and the analysis of "crossovers" and withdrawals from trials sometimes have important implications. In the trials of CABG, a large proportion of patients randomized to medical treatment eventually undergo surgery. These "crossovers" are so numerous that these trials do not compare surgical with medical treatments, but rather immediate surgery with initial medical treatment followed by surgery if indicated. That is to say, the trial tests a question of medical management rather than one of clinical efficacy. Data analysis in CABG trials is by "intention to treat." In some cases data are analyzed according to actual treatment, or the analysis may include both options.

Counting Events

Important methodological issues have been raised by a recent multicenter double-blind RCT, the Anturane Reinfarction Trial (ART). This RCT compared a placebo with Anturane (sulfinpyrazone), a platelet-active drug (one that inhibits blood clotting), in preventing cardiac mortality after myocardial infarction. A publication of the trial's results appeared in *The New England Journal of Medicine* in 1980 (4), reporting a reduction in cardiac mortality as a result of the drug. The difference was attributable to a decrease in sudden deaths (those deaths occurring within the first 6 months after myocardial infarction) in the experimental group. FDA later criticized the study on two grounds (220): 1) that the criteria used in classifying causes of death were ambiguous and illogical, and 2) that the criteria were not applied consistently. FDA also questioned the exclusion of certain participants and deaths in the analysis. Reanalysis of the data, including a reclassification of deaths by an independent group and by the ART Policy Committee, showed different results, though the same trend that was originally reported. The observed difference in overall mortality was no longer significant, though there were still fewer sudden deaths in the Anturane group compared to the group taking the placebo (3).

The disagreement over the ART in part concerns the way events are counted and attributed in RCTs (196). Decisions about which participants and events should and should not be counted in the analysis to some degree rest on whether the trial is considered one of medical management or clinical efficacy, though there is debate even on this point. In medical management trials all randomized patients are included in analysis, and all events during followup are counted. In trials of clinical efficacy, designed to test the biological effects of interventions, only those patients actually taking the treatments as prescribed are included in analysis, and only those prespecified events likely to be influenced by the treatment are counted. ART was a trial of clinical efficacy using debatable rules for counting, as well as some faulty applications of these rules.

Methods of Randomizing

Randomization does not ensure the equal distribution of characteristics, but it does ensure the valid use of statistical significance tests. Improper randomization, which has occurred many times, ensures neither. Various allocation schemes, more and less successful at randomization, have been based on date of birth, date of visit to the physician or hospital, alternating assignments as patients enter a trial, and other plans. Mosteller, Gilbert, and McPeek (165) review the biases of faulty allocation schemes. For example, in using the flip of a coin or the draw of a playing card, investigators might be tempted to even out groups if they begin to look unbalanced. Alternating assignments can be biased when two patients arrive simultaneously and a decision must be made about who gets which treatment. Physicians may know what the next treatment is and schedule patients accordingly, or they may selectively enter patients only when they approve the next "random" assignment.

In spite of such practical problems, random numbers can be reliably obtained from tables and from computer programs, and there are methods to ensure that investigators do not know which treatment a participant will be assigned. For example, in many multicenter trials treatments are assigned by telephone after patient eligibility has been established. The person enrolling a patient, therefore, has no control over group assignments.

Deviations From Treatments and Protocol

In the course of an RCT, events may not take place according to plan. In one well-known case, high-oxygen environments were evaluated as a possible cause of retrolental fibroplasia (a condition leading to blindness) in premature newborns. Some attending nurses in one of the studies were so strongly convinced that low-oxygen environments were harmful to the infants that they increased the levels of oxygen. Recognizing this practical problem in carrying out the trial, in another study the oxygen concentration was only partly reduced until the harmful effect of high oxygen concentration was firmly established (252). Not adhering to a protocol, as in the first study above, may invalidate the findings of an RCT if the deviation is widespread or unknown in extent. An investigator's lack of adherence to study protocol is probably the most serious type of deviation.

Patients may also deviate from the study protocol. In general, however, their lack of compliance, unlike that of investigators, can be planned for as another aspect of the RCT itself. Protocols can be designed to allow some patient noncompliance without compromising the results. RCT designers usually want to know about clinical efficacy in both experimental and ordinary conditions, making a certain amount of compliance necessary, on the one hand, and the quantifying of compliance necessary on the other. In some cases the percentage of compliant patients may be as important as the biological effect of the intervention, and compliance itself may be designated as an experimental endpoint. If a drug, for example, is known to be effective but patients will not take it, it has little value.

Blinding

"Blinding" is keeping secret the treatment assignments (experimental or control) of trial participants (see ch. 1 for more discussion of blinding). Blinding compensates for the expectations of patients and physicians which, whether positive or negative, can affect the experiment's outcome. A patient's sense of well-being may be enhanced by belief in a treatment, and a physician's assessment of the patient's condition may be strongly affected by the physician's expectations about the treatment.

Blinding in drug trials is accomplished commonly by the use of a placebo, usually an inert substance resembling the experimental drug. Blinding can fail even using a placebo, if, for example, the experimental drug has unmistakable side effects. A failure of blinding can raise doubts about an experiment's conclusions.

Blinding is not possible in some trials, notably those comparing surgical and medical treatments or other markedly different interventions. For example, in the Multiple Risk Factor Intervention Trial the experimental group received intensive counseling while controls went their normal route of care (166). The question arises in such a case whether the effects observed in experimental subjects are attributable to the treatment itself or to the attention they received. If all such trials are considered purely medical management trials, the importance of that distinction is diminished.

Other Issues Concerning the Quality of RCT Research

One criticism of most RCTs, which probably applies to much clinical research, is the information they fail to obtain on how interventions affect "quality of life." McPeek, Gilbert, and Mosteller (152) focused some attention on this issue based on a review of research evaluating new surgical procedures. Many RCTs show that as far as they can be measured, the interventions compared cannot be distinguished in efficacy or safety. Such is often the case in RCTs of cancer treatments. Thus, an important factor in deciding between therapies is the way they affect the patient's quality of life. Research in this area requires developing methods to define and appraise quality of life and developing administrative methods for the long-term followup of pertinent questions without great inconvenience to physicians and patients. Greater cooperation between social and clinical scientists has been recommended to develop RCTs (152).

Little is taught about clinical trials in medical schools, and from this might result poor quality of design and participation in RCTs. Improving physicians knowledge of the value of RCTs and of their conduct, both in medical schools and in continuing medical education, could motivate their better participation in RCTs.

Statistical Power and Statistical Significance*

A frequent criticism of RCTs is that they have lacked sufficient statistical power to detect important effects. In practical terms, this means that the number of cases studied is so small that even if the experimental technology is superior (or inferior) to the control treatment, the difference will likely not be detected in the RCT. Failure to detect such an effect is called a "Type II error," and is analogous to a "false negative." The probability of this type of error is expressed as "beta." "Power" is equal to 1-beta. Commonly sought power levels are 0.80 and 0.90.

Another type of error, less frequent in RCTs but closely related to lack of power, is concluding that there is an effect when, in fact, there is none. This can and does occur purely by chance because of sampling error. It can lead to adopting or rejecting a treatment that is neither more nor less effective than the tested alternative. This is known as "Type I error" and is analogous to a "false positive." The probability of this kind of error is expressed as "alpha," which is commonly called the level of statistical significance. Common alpha levels are 0.05 and 0.01.

The power of a trial is the probability of detecting an effect of at least a specified magnitude at a specified level of statistical significance. For example, a trial might have a power of 0.80 to detect a 50 percent better outcome in the experimental than in the control treatment at the 0.05 level of statistical significance.

As power is a function of sample size, it is essential in designing an RCT to determine the sample size needed for an effect of a specified magnitude to be judged statistically significant. Specifying the magnitude of effect depends in turn on the investigator's judgment of how large an effect would be practically significant and at the same time, how large an effect can be realistically expected. The larger the sample size, the higher the probability the test has of detecting an effect of a given magnitude, or, alternatively, the smaller the effect the test can detect as statistically significant. As sample size increases, however, so does the cost of the study. It would be wasteful to choose a sample size so large that it would detect a difference that has no practical significance. The investigator must make a judgment weighing cost and statistical power. Investigators frequently overestimate the effectiveness of the treatment under study and therefore underestimate the size of sample needed to detect a statistically significant effect. For example, the sample size may be chosen on the premise that the experimental treatment is 50 percent better than the control treatment, whereas in reality it is only 20 percent better. Statistical analysis is likely to lead to the erroneous conclusion that the experimental treatment is "not statistically significantly better" than the control even though the investigators might have considered the improvement of 20-percent important. Had the investigators chosen the larger sample size needed to detect a 20-percent improvement as statistically significant, they would have avoided this Type II error.

Small studies do have a place in the greater scheme of research, as pilot and feasibility tests, and, should a real breakthrough occur, they can detect such a big effect. Small studies in themselves are not the problem. Too often, though, they are treated as definitive, and not evaluated in light of their probability of finding a true difference.

Small study sizes and concomitant lack of statistical power are well illustrated by reviews of published cancer RCTs. Mosteller, Gilbert, and McPeek (165) surveyed the sample sizes in over 400 trials referred to in the volume *Randomized Trials in Cancer: A Critical Review by Sites* (211; discussed in ch. 5) as well as 54 RCTs from the journal *Cancer* that Zelen and colleagues review in an earlier paper (258). Zelen concluded that the median sample size was about 50 per treatment group. Mosteller and colleagues (165) found this calculation to be "a bit optimistic."

A "typical trial," conducted on 50 patients, has a probability of less than 0.40 to detect a difference from 20 percent of patients responding in the group, to 40 percent in the other (at the 0.05 level of significance) (258). Referring to these same data, Miké (161) noted that the studies could provide reasonable power only for differences in outcome so large as to be highly unlikely.

---

*This section benefitted considerably from reference 145.

Zelen (257) has addressed the problem of false positives in cancer research. Given the small sample sizes and the low probability of success of most trials of cancer therapies, Zelen calculates that of every five such trials with positive results, only two are true positives (see ch. 5 for a fuller discussion). Most positive results are published without being confirmed in a second trial, and oncologists are prone to accept them uncritically into their practices (258).

Haines evaluated the statistical power of published RCTs in neurosurgery. Of the 51 trials published since 1945, half had less than a 50/50 chance of finding a difference in outcome as large as 50 percent between the experimental and control groups (103).

Sometimes the sample sizes chosen for studies are based on unrealistic estimates of a treatment effect, making the studies too small to detect lesser but still important effects. Clinicians dream of spectacular new therapies, but in fact most progress occurs in small, incremental steps. Statisticians should be conservative in determining necessary sample sizes, and should aim for significance levels higher than are seemingly needed (25).

Greater cooperation between statisticians and clinicians is a way to improve the quality of trials. Haines showed that the best sign of a well-designed trial in neurosurgery was the participation of a statistician. This is probably true in a wide range of research. About 10 years ago, the National Center for Health Services Research studied the factors that affect approval of research grant applications. They found the most important single factor was the presence of a biostatistician on the proposed staff. Presumably this finding reflects the work of a biostatistician in preparing the proposals and accordingly, the proposal's substantive merit, rather than the mere presence of a biostatistician's name (145).

Recruitment

Many studies are never completed or not adequately completed because of poor patient recruitment (87). This stems, at least in part, from the tendency of clinicians to overestimate the number of patients available for study. Brown (25) states that clinicians overestimate the number of patients that can be recruited by at least twice, and sometimes as much as 10 times.

The problems of recruitment were graphically illustrated in the National Heart, Lung, and Blood Institute's (NHLBI) Lipid Research Clinics Coronary Primary Prevention Trial. The protocol called for 3,550 men to be recruited from physician referrals, advertisements in the media, clinical laboratories, blood banks, occupational screening, and other sources. The number of likely subjects was seriously overestimated, causing the project to fall behind schedule. While 46.5 percent of those referred from physicians and laboratories were recruited, only 2.5 percent of those from the other sources were. This experience was not unique. Tallying the numbers from four large-scale studies—this study, the National Diet Heart Study, part of the Hypertension Detection and Follow-Up Program, and VA's Mild Hypertension Study—almost 1 million contacts were screened to yield about 11,000 entrants (129).

Recruitment should take place as quickly as possible to avoid time-dependent trends that may complicate comparisons between patients recruited early and those recruited later.

The need to recruit many patients quickly has led to greater numbers of multicenter trials, an arrangement that appears to improve the quality of trials for reasons other than reliance on sheer numbers (see "Multicenter v. Single Center Trials," below). A related development, especially in RCTs of cancer treatments, is including community hospitals along with major research and teaching hospitals in multicenter RCTs. This reflects the trend of treating cancer patients in the community setting.

Multicenter v. Single Center Trials

More than half the RCTs in the 1979 NIH Inventory of Clinical Trials involved the participation of more than one institution. Such trials have a number of advantages.

Regardless of the experiment's protocol, recruiting at a number of institutions shortens the time necessary to enroll the participants. Such trials may take longer in planning, but prolonged recruitment can cause difficulties for RCTs (see "Recruitment"). In studying rare diseases, the cooperation of a number of centers is necessary to enroll even a modest number of patients. Permanently constituted "cooperative oncology groups" have been a mainstay of cancer therapy RCTs, especially in allowing clinical trials of therapies for rarer cancers (see ch. 5, "Impact of the Cooperative Oncology Groups"). The use of multiple centers has made possible the large-scale prevention trials in heart disease. Because of their larger sample sizes, multicenter studies generally have greater statistical power than single-center trials.

A second advantage of multicenter trials is that they often have more highly refined protocols and organization. In well-run trials, all investigators participate, both in planning and throughout the trial. Problems are likely to be worked out early. The effects observed in the trial are not likely to result from one investigator's personal style. Multicenter trials generally have better arrangements for data analysis and data monitoring, and more often employ statisticians in planning the collection and coordination of data.

A third advantage of multicenter RCTs is that they can enroll a more heterogeneous patient group. One criticism of RCTs, and a reason sometimes offered for the irrelevance of their findings, is that RCT participants represent only a small proportion of the total patient population. The results lack external validity, that is, they can't be generalized to real treatment decisions. Multicenter studies do not entirely eliminate this problem, but insofar as they are geographically distributed, the heterogeneity of the patient population is increased.

Traditionally, most institutions participating in multicenter RCTs have been large university research hospitals. (One exception is VA Cooperative Studies Program trials, carried out in VA hospitals.) More recent trials have sought to include community hospitals and small group practices, with varying degrees of success. One investigator claims that the data submitted by smaller institutions are inferior to those of the larger institutions (215). This claim has been questioned by multicenter research groups that include smaller institutions. They argue that in well-organized trials with strong central administration and sufficient training and orientation provided for the smaller institutions, no such difference can be seen (14). Thomas and colleagues (221) comment that "more clearly written protocols, orientation sessions for physicians, and more effective monitoring of satellite performance would go a long way toward keeping protocol studies open to a broader array of institutions, physicians and patients. This is particularly desirable if the knowledge gained from protocols is ever to be incorporated into standard treatment."

There are also arguments made against multicenter trials. For example, some argue that the complex administrative arrangements these trials require, if there is no established cooperative system, are too great an impediment. Multicenter trials are generally more expensive than single-center trials, mainly because of the number of participants. In fact, they are not necessarily more expensive per patient. Meinert calculated that the cost per patient in a multicenter RCT (based on the 1979 NIH Clinical Trials Inventory) is $523, while that for single center trials is $587 (158).

Even when multicenter trials are preferred in resolving clinical questions, there is a role for single-center investigations. First, there is a legitimate need for small-scale preliminary studies in the early stages of evaluation. Almost everyone would agree that RCTs should not be undertaken without some evidence from smaller studies on which to base the trial. In some cases these preliminary trials might be HCTs rather than RCTs. There are technical limitations to multicenter trials in that they require special skills or equipment. Unfortunately, multicenter trials may be foregone simply because the details of their design and execution are not the sufficiently known. In some poorly planned studies, data collection is expected to be part of regular patient care, and is not seen as research requiring extra time, an incorrect assumption.

Multicenter trials are often viewed as overly complex and not worth the effort. They are difficult to begin without some funding, and the initial stages of planning usually require more money than is available. As a result, the planning of large-scale trials in some fields falls more often to the Federal Government and not to other researchers in the field. This has been the case with NHLBI-funded trials, while the impetus for developing trials funded by the National Cancer Institute (NCI) is largely in the hands of the extramural community. Incentives to participate are less when investigators have little or no say in the design of trials.

Another problem of conducting multicenter trials is the lack of written material about the methods of large-scale RCTs, although this is changing. In the past few years, a number of articles have addressed such questions, including a number of articles in the journal *Controlled Clinical Trials.*

Investigators have little incentive to participate in multicenter RCTs because participating investigators are given little recognition. The "author" of publications reporting the trial is often given as the name of a group, e.g., the Multiple Risk Factor Intervention Trial Research Group (see ref. 166), and institutions award little status for participation. Academic promotions are rarely based on participation in large trials (159). (Related recommendations to encourage multicenter RCTs are discussed in ch. 6.)

Dissemination of RCT Results

As the number of trials conducted, including large-scale trials, has increased, their results are not so effectively disseminated by simply publishing them, even in distinguished journals.

The trials drug companies sponsor for FDA approval of New Drug Applications are often reported in obscure journals. Because drug companies have their livelihoods at stake, they take other steps toward disseminating their results. The two main avenues they use to reach the practicing physician are advertising (both in major journals and, perhaps more importantly, in "throwaway" publications) and the use of representatives who visit physicians' offices. The throwaway publications are distributed free of charge to most practicing physicians in the country. Advertising in major medical journals also receives widespread attention.

Drug companies' representatives, their salespeople, personally visit private physicians and medical institutions to distribute literature on their products, to dispense samples to physicians, and to encourage the physicians to prescribe their products. In general, neither advertising nor drug companies' representatives stress the design and conduct of the trials, but rather the uses of the drugs.

In a study of physicians' prescribing practices, Avorn found that "pharmaceutical advertising has become the major source of continuing education for the American physician" (156). This study indicated that both advertising and drug company representatives have a marked influence on prescribing habits, yet that most physicians believe both have only minimal influence.

The research community could profitably borrow from the practices of the drug industry in disseminating their results. It is very likely that research results would be better disseminated if increased resources were devoted to the effort. Funding bodies should recognize this more fully. At NIH, NHLBI, for instance, has a well-developed strategy toward disseminating research results (described in more detail in ch. 5). "Analysis and Dissemination" is a separate phase of all NHLBI's large-scale trials, and the Institute requires a plan for dissemination of trial results. The vehicles of communication it recommends are conferences, activities of professional societies, workshops, and articles in less specialized medical publications and the popular press. NHLBI's methods of dissemination are still evolving, but its progress is apparent. Its recently completed Multiple Risk Factor Intervention Trial received attention in all of the major medical journals, in newspapers and magazines, and on radio and television. NHLBI followed up the publications with a workshop (February 1983) to discuss the results of the trial with practitioners and policymakers. Other NIH institutes, such as the National Eye Institute and NCI, have also developed mechanisms to disseminate research results. While every trial cannot expect to become famous, efforts to publicize results, should be greater, including important negative results.

Effective dissemination of results depends on knowing how physicians get information. Medical journals and textbooks, continuing medical education courses, and discussions with colleagues appear to be the most influential sources aside from drug advertising (214). Depending on the subject, multiple sources of information may be important. Experimental programs have effectively used physician tutorials in hospitals for selected problems in the management of their patients (122). Nevertheless, not enough is known about how best to translate clinical research findings into practice.

At present, much dissemination of information is left to chance. Kessner has suggested a few measures to improve the situation:

1. identify the primary audience the results should reach,
2. communicate early with selected journal editors, and
3. allocate a small percentage of research funds to dissemination (122).

Other Factors Affecting the Impact of RCTs

Other characteristics of RCTs influence their impact. For example, investigators and their institutions, especially those of repute, can influence the acceptance of results.

Whether an RCT's results are negative or positive can affect its impact. Positive results are generally more enthusiastically embraced than negative ones. Positive results are also more likely to be published than negative results, and thus may have a greater impact.

The risk associated with a technology affects the way practitioners use information about its efficacy. Technologies perceived to be of low risk, such as many diagnostic tests, may still be used despite evidence questioning their efficacy. Some time-honored treatments, such as bed rest for hepatitis, persist despite the evidence, typifying the "it can't hurt" philosophy (40).

5.

The Impacts of Clinical Trials on Medical Practice

5.

The Impacts of Clinical Trials on Medical Practice

The use of randomized clinical trials (RCTs) grew enormously in the late 1960's and 1970's. By the mid-1970's, literature began to appear about their impact on medical practice. The interest in RCTs has continued to grow, but the body of literature evaluating their impacts is still small.

RCT results can have several effects. They can encourage the adoption or abandonment of technologies through treatment decisions by individual physicians and by institutions (e.g., those resulting in the purchase of equipment or in establishing screening programs), or through changes in policy, for example, Federal guidelines (e.g., for immunization practices). All these effects, insofar as they are actually supported by RCT results, are positive.

On the negative side, an RCT favoring the use of a therapeutic agent may encourage the agent's extensive but unjustified use. The drug cimetidine (Tagamet®), for example, was found in an RCT to be effective for treating duodenal ulcer. It then became widely used for conditions and indications for which it had never been tested by RCT (51).

RCTs are only one kind of research that can be done on a promising medical intervention, however. Because they are not the sole source of evidence, it is difficult to separate their impacts from those of the other factors.

The literature about the impact of RCTs is of two general types. The first begins with the results of specific RCTs or the results of RCTs in a specific area (e.g., RCTs of treatments for hypertension), and then examines whether physicians are aware of the results, or what their treatment practice is compared with the recommendations that arise from the RCTs. The second type starts with medical practice, either through literature reviews or by questionnaires, and determines how well practice agrees with the results of appropriate RCTs. An important element of some papers is their quantification of the delay between publication of RCT results and changes in practice. Many papers that describe RCTs and their results also make claims about their impact, but without citing supporting data. These papers are difficult to interpret.

An increasing number of papers review the results of a number of RCTs in a field and make recommendations for practice in light of those results. These range from qualitative reviews of the literature to formal statistical "meta analyses" synthesizing data from more than one study into a single set of statistics.

Most authors conclude that the impact of RCTs on medical practice has been less than optimal or that their impact is exceedingly slow to develop. The literature as a whole demonstrates great variation in the use of RCTs and in their influence in different medical areas. These studies of RCTs' effects have evolved in method. Earlier papers concentrated on showing the lack of influence of RCTs. More recent articles, going beyond simply showing this fact, have identified some of its possible explanations (discussed in detail in ch. 4). Information from all these studies has contributed to researchers' and funding agencies' greater awareness that the dissemination of research results plays a major role in determining their impact. The National Heart, Lung, and Blood Institute (NHLBI) is now taking more rigorous measures to disseminate the results of RCTs, and to make followup studies of how profoundly these results have affected practice. NHLBI has just completed a followup of two recent large-scale RCTs, the Coronary Drug Project (CDP) and the Aspirin Myocardial Infarction Study (AMIS), and plans similar followup of the recently completed Multiple Risk Factor Intervention Trial (MRFIT) and the ongoing Lipid Research Clinics. The National Cancer Institute (NCI) also has instituted a major new program for disseminating information about ongoing studies. Protocol Data Query System (PDQ) is an international computerized data base currently including information about treatment protocols for about 700 research programs.

The published literature on the impact of RCTs by no means covers all medical practice. More attention has been given to the impact of RCTs in cancer research, though there is now increasing interest in RCTs related to cardiovascular disease. These two medical areas have inspired the majority of clinical trials and the greatest expenditures for such trials. A 1981 conference on the recent history of RCTs, concerned at least in part with their impact, focused on cancer and heart disease. (The proceedings were published as the September 1982 issue of *Controlled Clinical Trials*.) A good deal has been written about RCTs in surgery. The main complaint for surgical RCTs is that too few are done, and that when they are done, they are late. Some authors have focused on controversial trials that illuminate particular issues, for instance, the University Group Diabetes Program (see box E in ch. 4), perhaps the most controversial trial of all time. The remaining published articles about the impact of RCTs are about diverse topics from nursing practices to pediatrics.

Because of the extent of related literature, the influence of RCTs on treatment of cardiovascular disease and cancer and on surgery are specifically discussed in later sections of this chapter.

RCTs AND CONCORDANCE WITH MEDICAL PRACTICE

In one of the earliest articles on the topic, Chalmers concluded that physicians' practice in the 1950's and 1960's was often at odds with data from RCTs (39). McGrady came to the same conclusion in a 1982 survey of family practitioners. Asked about their treatment of a variety of common problems, there was little concordance between their practice and the results of controlled trials (149).

Christensen, Juhl, and Tygstrup reviewed 65 RCTs on treatment of duodenal ulcer and compared the results to recommendations in medical textbooks. They found that RCTs had little influence on these recommendations (49). Tygstrup, Lachin, and Juhl (224) concluded that the results of RCTs have had little effect on gastroenterological therapy.

In a discussion of various types of research studies in ambulatory pediatrics, Hoekelman concluded that the results of RCTs had little influence on physicians' behavior (114).

Moskowitz, Sacks, and Chalmers reviewed RCTs of alcohol withdrawal treatment. They concluded that such treatment using drugs had been established as superior to that using only a placebo. They then polled physicians about their practices and examined review articles on alcohol withdrawal treatments. In this case, the authors found that practicing physicians were using the treatment that RCTs had shown to be effective before it had been recommended in review articles (163).

Baum and colleagues focused on RCTs' effects on later research, instead of their effects on practice. After surveying clinical trials of antibiotic prophylaxis in colon surgery, they concluded that the results published showing antibiotics superior to a placebo apparently had little effect on the design of later studies (12).

In a preliminary report, Boissel and colleagues conclude that the results of RCTs had no influence on the prescribing habits of French physicians for four classes of drugs—beta blockers, long-acting nitrates, clofibrate, and platelet antiaggregants (19).

Stross and Harlan found that only 28 percent of family physicians and 46 percent of internists were aware of the results of a major multicenter study using photocoagulation to treat diabetic retinopathy (Diabetic Retinopathy Study [DRS]), a year and a half after the study had been published (213). Their study shows that even the results of well-conducted large-scale studies must be brought explicitly to physicians' attention or these results will not affect practice. The DRS was reported in an ophthalmologic journal, not inappropriately, but leaving uninformed the general practice physicians who usually treat diabetics.

Medical practice might have benefited more from DRS had it been given greater coverage initially, e.g., as a report of a clinical advance, rather than one of the study itself, in a general medical journal with wide circulation.

Stross and Harlan also found that many who knew about DRS had learned about it from ophthalmologists or other colleagues, not from the medical literature. This argues for encouraging communication among physicians in local areas. Continuing medical education could also give greater emphasis to new findings in clinical research.

The National Institute of Mental Health (NIMH) of the Alcohol, Drug Abuse, and Mental Health Administration played a key role in evaluating hyperbaric oxygen treatment for cerebral dysfunction in the elderly and also in seeing that the evaluation had appropriate impact (see box F).

NIMH has continued to fund RCTs when promising but controversial treatments appear. As of 1980, in response to reports that schizophrenics can be treated with hemodialysis (244), NIMH funded three double-blind RCTs, two still under way. Carpenter and colleagues (36) have reported their finding from the study that is complete, a small study of 15 patients. They used a "cross-over design" for the study. They randomized patients to one treatment or the other initially, and switched to the other treatment midway through the trial. The experimental treatment was dialysis and the control treatment, sham dialysis. Carpenter and his colleagues found no difference between the effects of real and sham dialysis on the symptoms and behavior of schizophrenia. The results of this trial (along with the other two) may have a direct impact on practice, depending on coverage decisions for the procedure by Medicare. In response to a request for evaluation from the Health Care Financing Administration, the National Center for Health Care Technology found that the evidence for the procedure's safety and efficacy was inconclusive and recommended that it not be covered under Medicare (235). With evidence from the other RCTs, this initial decision may be either affirmed or overturned.

RCTs IN CANCER RESEARCH

Characteristics of Cancer RCTs

RCTs are employed in developing cancer drugs in "phase III" clinical testing. Preclinical tests identify potential anticancer agents, and then test them in rodents and larger mammals. Phase I clinical studies establish the tolerated dosages of the drugs and their toxicities and measure any therapeutic effects they have. Phase II trials evaluate drugs in treating specific kinds of tumors. In phase III trials, RCTs are used to compare a new treatment with whatever the standard treatment is at that time.

Anticancer drugs are generally very active compounds with marked toxicities, and the patient populations on which they are tested reflect their risks. In testing most other kinds of drugs, phase I studies are carried out on relatively healthy subjects, and only later studies on those with the conditions for which the drug is intended. In contrast, the first clinical studies of cancer drugs are carried out on those with very advanced cancers, who have not improved through any other treatment, and for whom there is little other hope. These clinical studies then progress, if the drug shows promise, to testing the drug on patients with early cancers who are more likely to benefit from therapy.

The earlier the stage of a cancer, and the greater the survival rate for that kind of cancer, the less acceptable is treating that cancer using a drug with known and unknown risks, and unknown value. This fact has affected the use of RCTs in cancer research. More RCTs have tested treatments of acute leukemias, for example, than of chronic leukemias, in part because the acute forms were rapidly fatal, and at least in acute lymphocytic leukemia (ALL), most victims were children. People with chronic leukemias can live for years, and those affected are usually older.

Clinical trials of cancer therapies can be somewhat more complex than clinical trials of therapies

Box F.—Hyperbaric Oxygen Treatment for Cognitive Deficits in the Elderly*

Considerable excitement arose in both scientific and lay communities over a 1969 article in the *New England Journal of Medicine,* reporting that repeated exposure to pure pressurized oxygen in a hyperbaric chamber enhanced the cognitive functioning of elderly male patients with organic brain syndrome (117). No effective treatment had been available before for the memory loss associated with brain changes due to arteriosclerotic disease or Alzheimer's disease. This finding by Jacobs and her associates was especially compelling because five of their control subjects exposed to an air mixture failed to show improvement initially, but did improve later when they were "crossed over" to the oxygen treatment. Perhaps 10 percent of those over 65 years of age are affected by cerebral dysfunction, and so the potential impact of this therapy was enormous.

Five other published studies confirmed Jacobs' observation (16,22,66,115,116), but only one used a control group. Two additional studies failed to replicate Jacobs' findings (95,222). One using 21 experimental subjects and 4 control subjects failed to note any significant differences between the experimental and control subjects (222).

One of the major problems in evaluating the efficacy of hyperbaric oxygen as a treatment was the paucity of studies that employed control subjects and the small number of control subjects in those that did. One reason for the investigators' reluctance to include control subjects was that the control condition was more dangerous than the experimental one. Experimental subjects breathed pure oxygen, control subjects an air mixture containing nitrogen, presenting some danger of the "bends" if care was not taken in timing decompression.

Because of the importance of Jacobs' results and the obvious need for their confirmation using a sufficient number of control subjects, in 1973 the Psychopharmacology Research Branch of NIMH and the New York Medical Center undertook a collaborative RCT of the treatment.

This study failed to confirm that oxygen administered under pressure improves cognitive functioning in the elderly. The study had also investigated whether some subgroups of patients might be especially aided by the treatment. Again, there was no evidence of differential treatment effects as a function of initial severity of illness, sex, or presumed evidence of cerebrovascular disease. Subjects in the study had well-documented evidence of memory problems but were still able to reside in the community and to respond meaningfully to intelligence, psychological, and psychometric tests. On the basis of the findings of Jacobs and others (117), many of these patients should have shown a favorable response to hyperbaric oxygen treatment, but this was not the case.

Jacobs' findings had been picked up early on by the news media, especially the more sensational press, and hyperbaric oxygen was widely touted as a cure for a variety of the infirmities of old age as well as for memory loss. A number of special centers in this country were already offering hyperbaric oxygen to treat memory loss in the elderly at substantial fees. At one, the fee was $5,000 for 15 days of treatment. The problem of the established use of this treatment was not easy to resolve. Scientific findings are generally not disseminated widely prior to their publication in a respected scientific journal, where the lag time between receipt of a manuscript and publication may run a year or more. To offset this delay, researchers decided to present the new findings at a meeting of the American Geriatric Society and to release a statement to the press once word was received that the paper had been accepted for publication (186).

Although publication of the study findings and dissemination of the results through the press and television did not completely eliminate the practice, the coverage did appear to dampen enthusiasm significantly. The study findings also had an effect on the policy of health insurance carriers and that of the Medicare program, which at one time had considered paying for the treatment. The insurance carriers and Medicare have since ruled that use of hyperbaric oxygen is not a medically accepted or effective treatment for cognitive deficits in the elderly, and they will not pay for it.

By identifying the need for an RCT, and acting quickly, NIMH halted the spread of an ineffective treatment. This case points out the importance of appropriately disseminating scientific findings. Information that promises relief to suffering individuals may be disseminated quickly and extensively—perhaps exceedingly so—when testing has been inadequate. In such cases, later valid findings must be given the widest and most rapid dissemination possible.

---

*Adapted from *Assessing the Efficacy and Safety of Medical Technologies* (225).

BEST COPY AVAILABLE

for other diseases. Four major types of treatment are now given to those with cancer: 1) surgery, 2) chemotherapy (treatment with drugs), 3) radiotherapy (treatment with ionizing radiation), and 4) biological response modification. The best therapies now available for most solid tumors* combine several of these treatments. Most RCTs have tested chemotherapies and more recently, types of biological response modification. Chemotherapy itself is not a simple treatment. Combinations of three or more drugs often provide the best results. The possible variations in chemotherapy, including dosages, timing of drug administration, and types of drugs, are almost limitless. The greatest limiting factor for such possible variations is probably the number of active anticancer drugs available; there are now about 20.

Most RCTs in cancer research are of chemotherapeutic agents. Surgery and radiotherapy have been tested far less often, in part because the first has been a mainstay of cancer treatment since the last century, and the second, since early in this century. The major developments in these therapies occurred before RCTs were in common use.

At least two volumes and a number of papers have addressed specifically the impact of RCTs on cancer therapies. *Randomized Trials in Cancer: A Critical Review by Sites* contains a number of papers by experts on all major anatomical sites of cancer and groups of these sites. These papers review the bases for treatment and the contribution of RCTs to current recommendations (211). *Methods and Impacts of Controlled Therapeutic Trials in Cancer* (5,37), published as part of a project of the International Union Against Cancer, reports on RCTs from their initiation to their conclusion, and determines the extent to which the results have altered therapeutic methods in subsequent years. A second part lists treatments available for specific cancers, including colorectal, bronchogenic, breast, melanoma, and osteosarcoma, and attempts to identify the roles of randomized and nonrandomized clinical trials in establishing their treatments.

Impact of the Cooperative Oncology Groups on RCTs

The mid-1950's saw the development of NCI "cooperative groups," to carry out multicenter studies in cancer treatment. These groups conducted the first RCTs in cancer research, studying treatment for childhood acute leukemia and for a variety of solid tumors. Fourteen groups are now active: five include multidisease, multiprotocol studies; six specialize by disease (e.g., National Wilms' Tumor Study Group and National Surgical Adjuvant Breast and Bowel Group); and three are "related resource groups" (Lymphoma Pathology Reference Center, Radiologic Physics Center and Cancer Clinical Investigations Coordinating Center) (59). Each group consists of 30 to 50 institutions (59), with more than 1,000 institutions participating altogether, including affiliates from 41 countries outside the United States. While these foreign affiliates are rarely funded, they find it important to participate (35). The cooperative groups are active in phase II as well as in phase III clinical trials (RCTs).

One of the main advantages of the cooperative groups is that they can recruit relatively large numbers of patients for trials in far shorter time than can single institutions. As is discussed below, small studies abound in the cancer treatment literature, more noticeably than any other field.

From the administrative necessities of large cooperative efforts the groups have developed well-formed organizations. Each has an elected chairman, an elected or appointed statistician, and several other elected and appointed positions and committees. The scientific sections of the groups vary, but include committees representing treatment modalities and specific diseases. Another important feature of the cooperative groups is that each has a statistical coordinating center. As in other areas, the presence of statistical expertise is a key factor in ensuring the high quality of RCTs.

The Cooperative Groups ensure a high quality of research by stringent internal review mechanisms, in addition to the usual external reviews of Government-supported research. Group members are evaluated at regular intervals on specific criteria related to the quality and productivity of trials (35). These evaluations can include auditing original clinical documents for accuracy of reporting (255).

---

*There are three main classes of neoplasms or cancers. Cancers of the epithelia, including the external epithelium (the skin and the lining of intestinal and respiratory tracts) and internal epithelia (the lining of various glands) are called carcinomas. Cancers of supportive tissues (e.g., bone, muscle, tendon, and cartilage) are called sarcomas. Carcinomas and sarcomas together are termed "solid tumors." Cancers of blood are called leukemias and those of the lymph tissues lymphomas.

The Cooperative Group members have traditionally been university hospitals or major treatment centers. Cancer patients are increasingly treated in community hospitals, however, as more oncologists are trained and enter the medical work force. The Cooperative Groups have thus recently arranged for community hospitals to participate in clinical trials. This should improve the efficiency of trials by extending the population from which patients are recruited, and improve the impact of trials by involving a greater number of oncologists and institutions. The Eastern Cooperative Oncology Group (ECOG) published their first evaluation of community hospital participation in their clinical trials. It indicated that the contribution of 112 community hospitals is equal in quality to that of the larger member institutions. Quality was measured by relative enrollment rates in trials, compliance with the protocol, and submission of data, as well as measures of outcome—e.g., survival and positive and toxic responses to treatment. (Community hospitals have shown similar performance in multicenter trials of heart disease (83).)

ECOG has found in addition, through a survey of affiliated hospitals, that while 16 percent of cancer patients were enrolled in a trial, a further 35 percent were treated in accordance with an experimental protocol.

Impact of RCTs on Cancer Treatment

RCTs have contributed to developing successful treatments for a number of cancers, e.g., those for ALL, Hodgkins' disease, and Wilm's tumor, and adjuvant chemotherapy for breast cancer. The clinical trials for these therapies have been part of larger targeted research programs, which were prompted by the discovery of significant drugs. The therapeutic regimens now actually employed were then developed gradually by trying the different drugs and their combinations in RCTs and building new trials on the results of previous ones. The sustained support of these programs and rational process through which they developed treatments appear to be the reasons for their success. Had uncoordinated trials been conducted in many places after the initial discoveries were made, it is doubtful that this progress could have been made as quickly and efficiently. It can be argued, on the other side, that new approaches and ideas may have been sacrificed by concentrating the effort.

RCTs have also had a major impact, though one difficult to document or quantify, in preventing costly but ineffective and debilitating cancer therapies from becoming part of medical practice (208).

Garnier, Flamant, and Fohanno (86) have shown that RCTs in cancer research are not conducted in proportion to the incidence or importance of the disease, but are heavily influenced by whether or not worthwhile treatments are available to be tested (table 6). While the highest incidence of cancer is at sites in the gastrointestinal tract, only 10.8 percent of RCTs are on treatments for cancers at those sites. The leukemias and hematosarcomas (circulatory cell neoplasms) account for 26.7 percent of RCTs, while the incidence of these cancers is less than one-third that of gastrointestinal cancers. The RCTs referred to here are those registered with the International Union Against Cancer between 1968 and 1978, nearly 1,000 RCTs.

A series of therapeutic advances, such as in treating ALL, depends on an initial breakthrough. For most cancers, particularly the solid tumors, such breakthroughs are rare. Most clinical trials in treatments of these tumors consist of testing drugs that have shown anticancer activity against a number of tumor types in phase I and phase II trials. These trials are usually small and conducted at single centers, with too few participants to showing a significant effect of the drug, if it has one. In part this is because a "significant" effect of an anticancer drug may be smaller than such an effect in treating less serious and more treatable diseases.

Thousands of cancer therapy RCTs have been generated by combining chemotherapeutics, often

Table 6.—Distribution by Site of the 945 Trials Registered at the International Union Against Cancer Information Office, and Related Incidence Rates

Site	Incidence[a]	Percent of trials by site
Gastrointestinal tract	76.0	10.8
Genito-urinary sites	46.9	5.9
Breast	40.4	15.9
Lung	40.0	12.5
Gynecological sites	30.4	6.9
Leukemias and haematosarcomas	23.0	26.7
Miscellaneous	15.3	3.3
Head and neck	15.0	5.4
Brain and nervous system	4.9	4.3
Skin (including melanoma)	4.2	5.4
Bone and soft tissue	2.8	3.0

[a] Average annual age-adjusted incidence rate per 100,000 population, United States.

SOURCE: H. Garnier, R. Flamant, and C. Fohanno, "Assessment of the Role of Randomized Clinical Trials in Establishing Treatment Policies," *Contr. Clin. Tr.* 3(3):227-234, September 1982.

two to four in one regimen, along with radiotherapy and surgery. Though drug combinations are based on some prior information, there is no satisfactory scientific basis for designing combinations. Given that the prior probability of success—the expectation that the trial will have positive results—is low in cancer research (judging from the history of cancer therapy RCTs), and that most of these RCTs employ few patients (a median of 25 per treatment), a large proportion of the positive results obtained must be false positives. The consequence is that many ineffective treatments may be applied in the clinic because clinicians do not have adequate information to distinguish effective from ineffective ones.

Many of the contributors to Staquet's book identified areas in which ongoing trials would provide some answers in the next few years and areas in which studies were needed (211). The contributors to the International Union Against Cancer's two-part publication concluded that RCTs have in most cases been more useful than nonrandomized studies in developing cancer treatments (5,37).

Garnier and colleagues looked at the treatment policies for head and neck cancers at the Gustave-Roussy Institute during two periods: from 1960 to 1967 and after 1967. They then examined the possible reasons for policy changes between the two periods. They set out to answer three questions about treatments for each main site of cancer: 1) whether there was a consensus about treatment, 2) the reasons for the choice of a specific treatment, and 3) the correlation between the treatment problems yet unsolved and the trials being conducted by the international cooperative groups (86). These authors did not complete the task they set for themselves. To have done so might have been a monumental undertaking. In fact, their attempt raises the larger question of how, whether, and to what end the impact of RCTs can be correctly and completely determined.

The authors did conclude, however, that there is consensus mainly about treatments that have not been tested in RCTs, namely those of surgery and radiotherapy.

Breast Cancer

The treatment of breast cancer has given rise to more RCTs than any other cancer site (37), and the impact of those trials has gradually been felt. In 1977, McPherson and Fox reviewed the reports of selected RCTs published since 1965, when the first RCT report demonstrated that radical mastectomy had no survival advantage over a more conservative operation (119). McPherson and Fox concluded that the RCTs had little impact: the radical procedure was still the treatment of choice based on surgery rates in 1970 (153).

A more recent paper on breast cancer (190) presents the view of the National Surgical Adjuvant Project for Breast and Bowel Cancers (NSABP), which is more optimistic about the impact of RCTs. Initial NSABP RCTs of breast cancer therapy focused on the treatment of local-regional disease (not metastatic), comparing radical mastectomy, total (simple) mastectomy with radiation, and total mastectomy alone with removal of axillary nodes only when they become affected. Underlying the study were competing hypotheses about the nature of breast cancer. The traditional belief, on which the rationale for the radical mastectomy is based, is that breast cancer follows an orderly progression from local-regional to systemic disease. The competing hypothesis is that the disease is often systemic very early on, so that even considerable improvements in local-regional treatment alone will not substantially affect the outcome of the disease. The trial results support the second hypothesis, with little difference in long-term survival observed among the treatment groups. The more extensive surgery involved in a radical mastectomy is not better than less extensive surgery in this regard.

The remaining NSABP trials have studied the effects of chemotherapy. Like the trials that developed a treatment regimen for ALL, these trials developed breast cancer treatments by increments. New trials are now being conducted in this area with a wide range of patients.

Continuing progress has been made through NSABP over the past 10 years, particularly in the use of adjuvant chemotherapy. The advances would have been difficult to document without the use of clinical trials in a structured program. In an overview of NSABP, the principal investigators of the program come to some generalizations about clinical trials of cancer treatments (190):

1. There is a need for larger sample sizes than are generally used in adjuvant phase III clinical studies. The heterogeneity of the patient population along a number of important prognostic lines, both known and unknown, make this particularly important.
2. Because of the relatively good prognosis for breast cancer patients, long followup is necessary, and overall survival, not necessarily disease-free survival, may be the appropriate measure.
3. The need for large numbers necessitates the need for multicenter participation. The development of straightforward, clear aims and reasonable data collection requirements is essential for success. In addition, particularly with long-term studies, constant refamiliarization of staff at participating institutions, where turnover may be high, is necessary.
4. Finally, the authors point to the need for clinical trials to be integrated into a general program aimed at the disease, which is predicated on an understanding of the natural history of the disease, and seeks to gain biological information about the disease.

The authors conclude that RCTs have contributed substantially to treating primary breast cancer in its early stages, and that NSABP trials have had a "strong impact in changing the clinical management of breast cancer over the past decade." Their conclusion is supported to some extent by trends in surgery for breast cancer between 1972 and 1981 (2). While the number of patients with breast cancer given radical mastectomies has dramatically declined (from about 50 percent in 1972 to about 3 percent in 1981), the shift has not been so much to simple (total) mastectomy or lesser surgery, but to a compromise between the radical and simple mastectomies, the modified radical mastectomy. In 1972, less than 30 percent of those with breast cancer had modified radical mastectomies; in 1981, over 70 percent. Between 1976 and 1981, there was a modest increase in women given a "wedge excision" (lumpectomy), from about 3 to 8 percent of those with breast cancer.

Early Detection in Cancer

The best secondary prevention for cancer is breast cancer screening. Miller and Bulbrook reviewed all major studies, randomized and nonrandomized, of all methods of breast cancer detection: self-examination, physical examination by medical personnel, thermography, mammography, and combinations of techniques. The combination of mammography and physical examination has proven most valuable (162).

The first trial of breast cancer screening, conducted by the Health Insurance Plan of New York, studied 62,000 women who were randomized either to mammography and clinical examination or to their regular pattern of care. The results showed a benefit of screening for women over 50 (204), though there is still some controversy over this study. Current studies in Canada and Sweden are designed to determine whether screening younger women is worthwhile (162).

Based on the available evidence, Miller and Bulbrook conclude that there is value in screening asymptomatic women over 50 by physical examination and mammography, but that the desirability of introducing screening on a larger scale requires answers to some outstanding questions. Studies in progress should provide the necessary information within the next decade. Regarding the potential impact of these studies on practice, "it should be noted that results from experimental studies cannot necessarily be directly translated into practice." This transition requires information in several areas: the training of personnel, the factors affecting participation in screening programs outside experimental settings, and the quality control of screening.

There has been relatively little improvement in survival for most common forms of cancer during the past three decades. Because survival is better for many cancers treated in earlier stages, early detection may hold the greatest current potential for lowering overall cancer mortality (226). Of such early detection techniques, breast cancer screening has received the most attention. There are also now three RCTs of lung cancer screening in progress, each testing both sputum cytology and X-rays. A preliminary finding in two of those is that sputum cytology is relatively ineffective. In addition, they have found that the benefits of screening, if proven, will be in detecting non-small-cell cancers* (85), which comprise the majority of lung cancers.

RCTs could also make the use of existing screening techniques more effective. The Pap smear, an examination of cells from the cervix, was introduced in 1943, to detect cervical cancer in asymptomatic women. The technique has been widely promoted and accepted, even though its efficacy has never been demonstrated in an RCT. In 1973, 75 percent of U.S. women over 17 had had at least one Pap smear. In recent years a controversy has developed about the efficacy of this screening, focusing on four issues: the natural course of cervical cancer, the accuracy of the test, the appropriate interval between screening tests, and the efficacy of screening while the incidence of death from cervical cancer is declining. OTA concluded (225):

> Once the Pap smear was in widespread use, the very extent of use and professional consensus of its efficacy argued against carrying out a controlled trial. As the risks to women whose tests were found falsely positive by the Pap smear have never been seriously documented, it is possible that a controlled trial to examine that question may be of value.

---

*Non-small-cell lung cancers include adenocarcinomas, squamous cell carcinomas, and large-cell carcinomas.

CARDIOVASCULAR DISEASE

The major problems in the treatment and prevention of cardiovascular disease have been well-studied in the United States, Canada, Europe, and Australia. RCTs are the primary instruments for resolving issues of therapy and prevention. NHLBI and the Veterans Administration (VA) have been key players in this field in the United States. Their large-scale multicenter RCTs, many with thousands of participants, have had a major impact on the treatment of heart disease.

These trials are mostly of two types: prevention trials based on evidence from epidemiology and physiology, and trials of therapeutic surgery and drugs. In the first category, the most intensively studied interventions for cardiovascular disease are those for lowering blood pressure, those for lowering levels of blood lipids and those for preventing thrombosis (blood clots), each of which has spawned large-scale primary and secondary prevention trials. Therapeutic trials have focused on surgical procedures (most importantly coronary artery bypass surgery), on beta-blocking drugs, and on antithrombotic agents. In general, trials for cardiovascular disease have not been undertaken without strong hypotheses to test and unless the intervention they test has a reasonably good chance of success.

RCTs of treatments for cardiovascular disease have progressed along a number of lines. One important trend in this field has been toward large multicenter trials. A second trend, illustrated by RCTs in hypertension, is a progression from those of treatments toward those of secondary, and more recently, primary prevention. The first major trials in hypertension studied severe hypertensives, and then later those with moderate and mild hypertension. A new NHLBI trial is testing interventions to prevent hypertension in those who are likely to develop it.

A third trend in research on cardiovascular disease results from knowing that it may have many causes. Early trials in the area concentrated on interventions related to single risk factors. More recent trials have studied several risk factors at once, notably MRFIT, which focused simultaneously on the risks of hypertension, high blood lipid levels, and cigarette smoking.

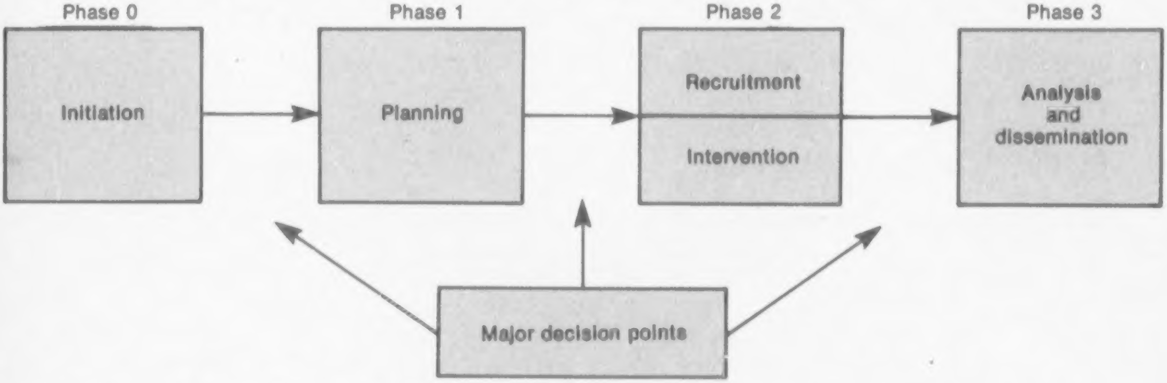
NHLBI and RCTs

NHLBI bases its decisionmaking about RCTs on an idealized view of the progression from basic research to health practice (fig. 1). The philosophy underlying NHLBI's use of clinical trials is well articulated by Levy and Sondik (134):

> Advances in knowledge at the basic research level result in hypotheses on potentially effective approaches for the prevention, management and control of disease in man. One objective of clinical research involves the testing of these hypotheses in controlled settings. Clinical trials serve to bridge clinical research and demonstration, prevention, education, and control activities. The clinical trial tests and validates the effectiveness of therapies before their introduction into the health care system. In some cases, however, trials are used to determine which of several alternative treatments already in use is most effective.

NHLBI's model could serve in other circumstances as one for decisions about clinical trials (fig. 1). Of particular relevance to this paper is NHLBI's phase 3, "Analysis and Dissemination." The success of preceding phases is, of course, required for that of phase 3: the initial concept must address an important question that can be answered in a clinical trial, planning must be ade-

Figure 1.—The National Heart, Lung, and Blood Institute's Clinical Trial Decision Process



SOURCE: R. Levy and E. Sondik, "Decision-making in Planning Large-Scale Comparative Studies," *N.Y. Acad. Sci.* 304:441-457, 1978.

BEST COPY AVAILABLE

quate to ensure answering the question, the trial must be carried out in accordance with the protocol and its progress well monitored. The dissemination of results depends on a well-designed, well-executed trial if the results are to have a positive impact on health care.

Data analysis is an ongoing activity in clinical trials, and interim results are sometimes published. The major effort to disseminate results follows the final data analysis, and begins with their publication in the medical literature. This is also their final resting place in many cases. NHLBI stresses that every institute-supported clinical trial must employ all available avenues of dissemination to be useful, including conferences, professional societies, workshops, and articles in less specialized medical publications and the popular press. A few months after the MRFIT results were published, for example, NIH held a 2-day workshop to discuss the results and their implications.

In addition to publicizing trials, it may be useful to find out how effective dissemination has been. NHLBI has completed a followup of CDP and AMIS (described below), and has a similar contract for the MRFIT and the Lipid Research Clinics.

The Coronary Drug Project and Aspirin Myocardial Infarction Study Followups

The fact that trials are well designed and well run does not guarantee that their results will influence practice. Given its heavy investments in clinical trials, NHLBI has an equal interest in knowing how influential they are. A few years ago NHLBI began an effort to find out the impact of two major RCTs, the CDP, which began in 1974 and AMIS, which began in 1980. It interviewed about 1,800 physicians nationwide about their knowledge of the studies and the studies' results, and about their treatment practices. Of all groups, cardiologists were the best informed, though probably not from having read the original reports of the trials. Internists and general practitioners were less well informed.

The results of the followups have not yet been published except in abstract form, and NHLBI has made no formal changes in policy for disseminating results, but the study suggests certain improvements. The dissemination of information must be local to reach most physicians. The national meetings of specialty societies already disseminate study results and treatment recommendations, but they could increase these efforts. Greater coverage of study results in the throwaway journals with wide circulations would reach physicians who don't read technical journals regularly.

RCTs and their impact on those areas of cardiovascular disease most actively investigated are described briefly.

Hypertension

High blood pressure, or hypertension, is one of the principal conditions leading to heart disease and stroke. The main strategies for controlling hypertension include diet modification, weight loss, behavior modification to reduce stress, and drug treatment. RCTs have tested several interventions in these areas, especially drug treatments.

Drugs to control hypertension first became available in the early 1960's following a search beginning after World War II. Their availability set the stage for large-scale RCTs. The VA Cooperative Studies Program (CSP) carried out the first large-scale RCT of drug treatment of severe hypertension (diastolic blood pressure [DBP] defined as above 115mmHg). The report of the study's results in 1967 showed convincingly that drug treatment helped to prevent death and disability from stroke, congestive heart failure, and kidney disease. A second study, published in 1970, extended the population studied to include men with DBP of 105 and above. Since that time, further studies in this country, under the auspices of VA and NHLBI, and in Europe and in Australia, have attempted to determine whether treatment of mild hypertension (usually defined as DBP between 90 or 95 and 104 or 109) also reduces morbidity and mortality.

Whether mild hypertensives should be treated with drugs is a question of more than passing interest. Perhaps 15 percent of the U.S. population has a DBP reading into the range of 90 to 104

DBP. McAlister describes this question as one " . . . with awesome social and economic implications" (146). Freis estimates that if the 40 million people in this country with blood pressures of 90 to 99 DBP all were given drug therapy, the annual cost of treatment might be as high as $20 billion (81).

In considering whether mild hypertensives should be treated, another important point should be weighed. There are qualitative as well as quantitative differences in the medical characteristics and of mild and severe hypertension. These affect the design of RCTs as well as the hopes for these patients' treatment. Severe hypertension has its own symptoms, in addition to its association with complicating disease. The treatment of severe hypertension both relieves these symptoms and reduces the risk of complicating disease. In contrast, mild hypertension is a symptomless condition. The major complication of mild hypertension is coronary heart disease. The major complications of moderate and severe hypertension are hemorrhagic stroke, renal failure, congestive heart failure, and aortic dissection (81).

With the trend toward treating milder hypertension in RCTs came the need for larger trials and proportional increases in cost. These trials illustrate a general point. The statistical power of trials (ch. 4, "Statistical Power and Statistical Significance") depend much more on the number of *endpoints* counted in each group than on the number of *participants* in a trial. Endpoints of importance in hypertension trials—stroke, heart failure, or death from some cardiovascular cause—occur much less frequently among those with mild than those with severe hypertension. Far more participants have been required for the later trials than those required for trials that tested treatments for severe hypertension. The first VA trial, whose participants were men with DBP over 115, provided convincing support for treatment with only 143 participants. The more recent Hypertension Detection and Followup Program (HDFP) required nearly 11,000 participants (about 8,000 with mild hypertension), and MRFIT, nearly 13,000 (about 8,000 with mild hypertension) for what was considered sufficient power.

The HDFP and MRFIT, along with a large Australian study (of about 3,400 with mild hypertension) and at least three smaller RCTs, have increased the debate over drug treatment of mild hypertension. All have provided information, but none an answer. The controversy focuses on the benefits of treatment and especially on the risks, known and unknown, of possible lifetime administration of antihypertensive drugs.

The HDFP showed that treatment reduced mortality by 20 percent in mild hypertensives (see box G). Pickering (183) puts this figure in a different light by expressing the 20-percent reduction in other terms, i.e., the reduction in the mortality rate from 7.7 percent in the control group to 6.4 percent in the treated group. In other words, of every 100 untreated patients, 7.7 died, while of every 100 treated patients, 6.4 died. Only 1.3 treated patients per 100 enjoyed a benefit. Pharmaceutical companies have used this information to claim that "HDFP findings justify early and aggressive management of mild hypertension," while some researchers have concluded that the studies provide no such basis for treatment (121).

The MRFIT study participants all had a high risk of cardiovascular disease, as defined by a rating included two other risk factors as well as hypertension: smoking and high blood lipid levels. A disturbing and unexpected finding in the MRFIT was a higher rate of death from coronary heart disease in the experimental than in the control group, in those hypertensive men who had abnormal baseline resting electrocardiograms. Subgroup analyses must be viewed cautiously, however, especially when they are not based on prior hypotheses. Nevertheless, in an editorial accompanying the MRFIT report, Lundberg commented that this result was "so major as to demand caution, since the results fly in the face of current medical dogma and practice" (138). His prediction that the observation would "no doubt foster substantial debate" was certainly correct. Only a few months after publishing the initial MRFIT results, the *Journal of the American Medical Association* carried two related articles and an editorial about the treatment of mild hypertension (121,146,183). Another related article, "Mild Hypertension: The Gray Zone Gets More Confusing" appeared in *Medical World News* during that interval (144). MRFIT results and resulting controversy have been publicized widely in both

Box G.—The Hypertension Detection and Followup (HDFP) Program*

The HDFP was a community-based RCT that studied 10,940 people with high blood pressure. The trial compared the effects on 5-year mortality of a systematic antihypertensive treatment program (stepped care, or SC) and referral to community medical care (referred care, or RC). SC patients were offered therapy in special centers, and therapy was increased stepwise to achieve and monitor reduction of blood pressure to specified levels. RC patients were sent to their usual sources of care, with special referrals for those with more severe hypertension or organ system damage. Patients were first grouped by age, sex, and race, and then further by the value of their DBP: 90 to 104; 105 to 114; and 115 or greater.

The study was designed to answer questions unresolved by previous studies conducted in VA's medical care system:

1. Is a systematic approach to antihypertensive therapy (SC) more effective in reducing risk of 5-year mortality for all hypertensive adults in the community compared to community care (RC)?
2. Can a substantial proportion of all hypertensives, detected in general populations, be pharmacologically managed to maintain blood pressure at normotensive levels?
3. Do the benefits of therapy exceed its toxicity in those with mild hypertension as well as in those with more severe hypertension?
4. Is antihypertensive therapy effective in young adults and in women and equally effective in blacks and whites?
5. Can morbidity and mortality from coronary artery disease be decreased by antihypertensive therapy?

The results of this large clinical trial, which cost nearly $70 million, showed that more intensive care with available therapies could lead to a significant decrease in mortality and morbidity from hypertension and that these benefits were found in treating "mild" hypertensives as well.

The results of HDFP were first published in the *Journal of the American Medical Association* in December 1979. A survey of physicians revealed that 40 percent of family physicians knew of the study within 2 months of publication, and 63 percent of internists within 6 months. Of the family physicians who knew of the study, 98 percent were able to correctly answer questions about the observed reduction in mortality and the benefits of treating mild hypertension. Eighty percent of the family physicians and 50 percent of the internists learned of the study from medical journals, and 40 percent of the internists learned of it from continuing medical education courses (the remainder learned of the study from colleagues or the lay press).

In sum, as a result of these RCTs and related educational activities, the public is much more aware that hypertension is a disease with serious but preventable consequences. The new information developed in HDFP disseminated rapidly to the medical community.

---

*Based largely on *Technology Transfer at the National Institutes of Health* (235).

the medical and the popular press. Each of the major trials has contributed to knowledge of hypertension, but at such expense that some find the results disappointing.

The newest NHLBI supported trial in this area is one of primary prevention of hypertension through dietary interventions in those aged 18 to 40. These interventions include altering the intake of sodium and potassium and helping patients to lose weight. This represents a logical step in the progression of related drug and diet trials that have been completed. Medical researchers would like to reduce the need for drugs in treating hypertension. The drugs carry some risk and are expensive. In treating a younger population, these RCTs also move toward the goal of primary prevention.

From the beginning, the trials of hypertension treatments have had a major effect on medical practice and on the design of subsequent trials.

Part of the NHLBI strategy has been the National High Blood Pressure Education Program, begun in 1972 to educate the medical community and the public about hypertension. Surveys of public knowledge about high blood pressure conducted in 1973 and in 1979 showed the following changes. First, those believing that hypertension is a serious condition increased from 63 percent in the 1973 survey to 73 percent in 1979. Second, 83 percent of those surveyed in 1979 had had their blood pressure measured within the past year, compared with 73 percent in the 1973 survey. Third, about twice as many people knew in 1979 what constituted normal blood pressure. Fourth, 40 percent more people understood in 1979 that hypertension did not have reliable symptoms. And fifth, in the 1979 survey, more people knew that effective treatment was available, and more were also following their prescribed therapies.

The early VA studies provided the first clear evidence of the benefit of drug treatment for severe and moderately severe hypertension. The first evidence from RCTs on the treatment of mild hypertension came in 1979 with publication of the HDFP (see box G). Even before that time, 92 percent of New York State physicians who responded to a questionnaire were treating patients with DBP in the range 90 to 104 (121). Since the publication of HDFP and the results of a large Australian trial, the use of drugs in treating hypertension has probably increased (121). MRFIT results pointed out the need to reexamine treatment policies, which, as described above, are now being debated in the literature.

The progression of hypertension trials has been orderly. New trials have built on the results of previous ones, not only those carried out in this country by VA and NHLBI, but also on those of trials in other countries. The available data allow some conclusions to be drawn and the reshaping of questions that remain for this field of research. Pickering makes three summary statements about treating mild hypertension (183):

> 1. Cardiovascular risk factors other than BP [blood pressure] should be taken into consideration. Therapeutic benefit is less likely to be seen in patients who have a low overall level of risk than in high-risk groups. Thus, two groups who have so far shown no benefit (in both the HDFP and Australian trial) are white women and men younger than 50 years. There is, therefore, no sound justification to treat all such patients.
> 2. For those who are at relatively high risk, treatment is more likely to confer protection against cerebrovascular events than coronary heart disease.
> 3. In doubtful cases, there is nothing to be lost by delaying the start of drug treatment. In both the HDFP and Australian trial, there was a substantial decline of BP in the control groups during the period of observation.

Freis makes similar recommendations based on RCT results: "By such a discriminative approach, many millions of people could be spared needless lifelong exposure to drugs" (81).

The evidence from RCTs in this field "does not support dogmatic guidelines" (146), but they do provide physicians useful information in considering each patient individually. Rather than supplanting clinical judgment in treating hypertension, the results of RCTs would appear to enhance it.

Hyperlipidemia

Known from epidemiologic studies, the strong relationship between high blood lipid levels (cholesterol and other fats) and the increased risk of atherosclerosis, has led to many large RCTs aimed at lowering blood lipid levels in the hope of reducing death rates. One of the first of these trials was conducted in Norway from 1956 to 1963. Since that time, trials have been under way continuously, each building on the results of earlier trials. (Buchwald, Fitch, and Moore discuss the major trials in this field (26).)

A notable evolution has occurred in trials that study the lowering of blood lipid levels. Early trials tested dietary interventions. These were mainly secondary prevention trials, and included only individuals with proven atherosclerotic disease. Lowering saturated fat was accomplished either by controlling total fat intake, or by substituting unsaturated (e.g., corn or soybean oil) for saturated fat (e.g., animal fat and butter).

Around the mid-1960's, more emphasis was placed on lowering lipid levels with drugs, while dietary recommendations were often provided to both experimental and control groups. A number of large trials in the United States and Europe tested the most promising drug at that time, clofibrate. Early results of these trials were also promising (26). In later trials, however, notably CDP funded by NHLBI, the benefits of clofibrate were small, particularly in light of some serious side effects. A European primary prevention trial confirmed the risks of the drug. The use of clofibrate has declined since the results of these studies were published (82).

Clofibrate was one of five treatments tested in CDP. Of the remaining four treatments, three were discontinued before completion of the trial because of adverse, at times lethal, effects. The discontinued drugs were estrogen (given in two dosage regimens) and dextrothyroxine. The last drug, niacin, also appeared to cause unwanted effects. It was, perhaps, effective in preventing recurrent nonfatal myocardial infarction, but not in altering mortality rates.

The Lipid Research Clinics, a primary prevention trial, is using a cholesterol-lowering diet for all participants and the drug cholestyramine for the experimental group. Results from this study are expected by the end of 1983.

One RCT still under way has been relatively successful in lowering blood lipids, the Program on the Surgical Control of the Hyperlipidemias (POSCH). POSCH also uses the most drastic intervention for such control: partial ileal bypass to reduce circulating blood cholesterol levels. Survivors of one myocardial infarction with high serum cholesterol levels, but with no other major risk factors, are eligible for the trial. Not surprisingly, recruitment for this trial has been slow. Complete recruitment of the 500 subjects required for each group may not be achieved. Early results show a 31-percent reduction in serum cholesterol in the surgical group over the first 3 years. Even if successful, because this procedure is radical, and has significant though not yet fully known side effects, it is unlikely to become a model for secondary prevention of cardiovascular disease.

A recent generation of trials, notably MRFIT in this country and the Oslo Heart Study in Norway, are primary prevention trials that use modifications in diet as the intervention to lower blood lipid levels. Both trials include interventions for more than one factor related to cardiovascular disease.

For the most part, the results from lipid-lowering trials have been less than promising (26):

> All completed randomized clinical trials of lipid intervention for atherosclerotic cardiovascular disease have shown no convincing evidence for disease retardation, arrest, or reversal associated with plasma cholesterol reduction; albeit in no trial has cholesterol reduction been marked and in many it has been minuscule.

These trials have served important purposes, in spite of their disappointing results. First, they have provided evidence against a number of drugs that might have been widely used without the trials. In addition, all the major diet intervention trials have shown some therapeutic benefit, if not as much as hoped. The trials, especially CDP, have generated a great deal of information about the natural history of cardiovascular disease. One finding is that serum cholesterol does not appear to be as prognostically important after myocardial infarction as before. This finding has important implications for treatments following myocardial infarction and for RCTs conducted of those treatments.

Coronary Artery Disease

Early surgical RCTs for coronary artery disease tested a procedure called internal mammary artery ligation. The procedure was based on the hypothesis that if the mammary arteries were tied off, blood flow to the heart would increase. The technique, though never widespread, gained brief popularity in the 1950's. At that time, two RCTs were conducted, comparing this surgery with a sham surgery. (These are the only RCTs that have used a sham surgical procedure (251).) The studies showed the sham procedure to be "at least as effective as internal mammary artery ligation" in treating angina pectoris. The procedure was rapidly abandoned after publication of the RCT's results. Fisher and Kennedy attribute this rapid change to the RCTs themselves (74).

The surgery in this field now under study is coronary artery bypass graft (CABG) surgery. Over

100,000 of these operations are now performed yearly in the United States (74), having rapidly increased from their first use in 1968. CABG surgery clearly relieves the pain of angina pectoris, and this is the reason for its widespread acceptance. However, the use of the procedure appears to have gone beyond its accepted indications (235).

The debate over CABG, which has inspired both U.S. and international RCTs, is over whether the procedure prolongs life, and if so, in which subset of patients. Controversy arose when the initial results of the full CABG study were released in 1977 showing no difference in survival between medically and surgically treated patients. The *New England Journal of Medicine* ran an editorial by Hiatt decrying the haphazardness of assessing surgical procedures, and suggesting that more orderly tests were called for (109). The trial was scrutinized from all angles and criticized on a number of points, especially the high rate of mortality in the surgery group early in the study.

Fisher and Kennedy conclude that in spite of this controversy the VA study convinced some that, while CABG prolonged the survival of those with left main artery disease, its effect on the survival of other patients was equivocal (74). More recent data from the study have also shown significantly increased survival in patients with three-vessel disease (without left main disease).

Wortman and Yeaton have identified nine RCTs of CABG surgery since 1974 (253). The first RCT of CABG surgery to have a major impact was the VA Cooperative Study. Fisher and Kennedy claim that this study "has had the most impact among the randomized studies published" (74). The trial began as one of a different operation, the Vineberg Implant, in 1968. This procedure was changed to CABG when it became evident that CABG was a superior operation. The early results on CABG showed it was better than medical therapy in prolonging life for those patients with left main artery disease. These results were readily accepted.

After 5 to 8 years of followup, a European RCT of CABG surgery found significantly increased survival in patients with three-vessel disease, those with stenosis in the proximal third of the left anterior descending artery, and insignificantly decreased survival in patients with left main artery disease (69). This trial has not elicited the reaction that the initial VA results did, probably in part because it justifies practices already current.

An NHLBI trial scheduled to end in 1983, the Coronary Artery Surgery Study, has suffered from entering the game rather late. A number of centers would not randomize patients because the evidence from other studies favored surgical treatment. A large registry is being kept as part of the study, including patients at one of those centers not randomizing.

Fisher and Kennedy drew several conclusions from their review of surgical trials for coronary artery disease (74). First, they found that these RCTs, especially the large, multicenter trials, have had a significant impact on clinical practice. The influence has not been uniform, however, nor has it been associated only with the quality of studies. Results that agree with current practice are readily accepted, as was VA's first report that patients with left main disease benefit from surgery. Results at odds with practice, on the other hand, are carefully scrutinized and criticized (see ch. 4, "Constituency Behind the Intervention").

Wortman and Yeaton compared the results of randomized and nonrandomized studies of CABG surgery, and synthesized the RCTs' results (253). They point out the value of RCTs by showing that nonrandomized studies consistently overestimate the benefit of surgery compared with randomized studies. This conclusion held regardless of whether the endpoint measured was mortality, survival, or size of effect. The discrepancy could not be explained by differences in distribution of patients' risk categories, crossover rates, or the timing of the trials. The different results between the two types of studies occur primarily because nonrandomized studies find that the medically treated group fares considerably worse than RCTs find. The surgically treated groups were not so different in outcome, though their results were slightly better in RCTs.

Antithrombosis Trials

Blood platelet aggregation is an important factor in thrombosis and in atherogenesis. A number of agents have been tested to prevent this aggregation. Aspirin, a well-known inhibitor of platelet aggregation, has been tested on heart attack survivors in at least six RCTs. The NHLBI AMIS, the largest RCT in this field with over 4,500 patients, showed that aspirin had no effect on survival.

Soon after publication of this trial's results, the Society for Clinical Trials reviewed it along with five other studies (including two other newly published trials). Together these trials studied over 10,000 myocardial infarction patients randomized between aspirin and double-blind placebo controls. During the studies, 1,000 of the patients died. Each study individually provided no clear evidence of aspirin's benefit. Taken together, however, they indicated that aspirin did reduce the risk of death, though at a lower rate than the individual tests could reliably detect. It was estimated that the overall reduction in the odds of reinfarction in all six trials was 21 percent (standard error ± 5 percent) and that about 70 deaths had been prevented (126a).

Reviewing the evidence from the six aspirin trials, an editorial in *The Lancet* concluded:

> It may be that the small benefit indicated thus far by both the antiplatelet and the anticoagulant randomized trials realistically represents all that can be achieved by any form of interference with haemostasis in the months or years after MI [myocardial infarction].

Other antiplatelet agents have been evaluated in RCTs—e.g., Persantine (dipyridamole) and Anturane (sulfinpyrazone) (see ch. 4 "the Anturane Reinfarction Trial").

NHLBI is now funding jointly with NCI a primary prevention trial to test the hypothesis that aspirin may help prevent initial MI. More than 20,000 healthy male U.S. physicians have been enrolled as participants in a double-blind placebo-controlled trial of aspirin to prevent cardiovascular disease in addition to testing beta carotene (a precursor of vitamin A) for cancer prevention.

Beta Blockers

In 1965, a nonrandomized study showed a reduction in mortality in those given propranolol, a beta-blocking drug (106), after a myocardial infarction. Though beta-blockers clearly have antihypertensive, antiarrhythmic, and antiplatelet properties, the mechanism through which they reduce mortality after MI unclear. Nonetheless, since then at least 41 placebo-controlled RCTs have tested at least 7 beta blockers in varying regimens (128).

Completed trials have most reliably evaluated the effect of "moderately prolonged beta-blockade in the period after discharge from hospital" (128). While most of these trials were too small to demonstrate a statistically significant benefit (using $p = 0.05$), in nearly all the trials mortality was reduced in those who took beta blockers. When the trials are pooled, a strongly significant result emerges. Based on the joint results, the total number of deaths was reduced by about 25 percent in those who took beta blockers over the course of the trials. "This effect will be widely regarded as sufficient to justify routine use of long-term beta-blockade in many patients for perhaps the first year or so after discharge from hospital" (128).

It is gratifying that RCTs have produced reliable information in this field, but questionable whether so many trials were necessary. Rose comments that given limited resources, "this sort of uncoordinated proliferation has been extremely wasteful" (193).

Two big questions remain about treatment regimens for beta blockers: 1) whether treatment should begin "early" (between a few hours and about 3 days after the infarct) or "late" (3 days later or more), and 2) how long the treatment should last. A number of studies of early beta-blockade are in progress, and answers to these questions may be available within the next few years. It is generally thought that beta blockers are used extensively for treating heart attack patients, and that their widespread use preceded convincing evidence from RCTs.

SURGERY

The impact of RCTs on surgery has been minimal, largely because RCTs in surgery are the exception rather than the rule. When RCTs are done, they are often criticized for coming too early or too late in the life of the innovation (see ch. 4, "Timing of RCTs").

It is instructive to consider the origins of surgery. Most current surgical practice has its beginnings before RCTs were available as a tool—i.e., before the middle of this century. Historically, much of the practice of surgery was in setting bones or suturing wounds. These procedures are clearly effective. As in treating acute diseases, a surgeon would know quite quickly whether the treatment worked. In many cases, the treatment could be repeated (e.g., a bone reset) if it failed the first time.

The removal of diseased or cancerous organs also seems to make such good sense intuitively that the value of such procedures was rarely questioned. If the patient died, it was not necessarily a failure of the operation, but a sign that the patient was beyond help. The theory behind much cancer surgery, which has been available since the last century, is that survival depends on removing all diseased tissue. (This assumes that all disease is visible, and that no spread of cancerous cells in the bloodstream occurs until late in the disease. The treatment of breast cancer has shown this not to be the case.) Successful surgery, meaning an aseptic operation that the patient survives, was considered successful treatment, and for many operations this is a good rule. Long-term outcomes have generally not been considered.

The nature of surgical procedures contributes to the difficulty of testing them through RCTs. Bonchek compares RCTs for surgery to those for drugs (20). Unlike drugs, which are fixed compounds, surgical procedures evolve. The efficacy of a drug is in many ways unrelated to the skill of the physician administering it. In surgery, the skill of the surgeon is vital, and this skill itself changes over time. Love observes (137):

> Drugs come as packaged preparations to be given by dosage. Operations are conceptual plans that require execution, and the details of a given operation change with time among surgeons and from patient to patient. It should be abundantly clear that techniques for evaluating the one cannot be used to evaluate the other.

Bunker and colleagues attribute the limited use of RCTs in surgery to the "very real conceptual, practical, ethical, and economic difficulties of carrying out in adequate numbers and sizes experiments involving complex surgical procedures in human beings" (30). They also conclude that not conducting such trials can cost more in dollars and lives than a trial adequate to answer the question.

Surgical RCTs in cancer treatment follow much the same pattern as those in other fields. Trials of chemotherapy by far outnumber those in surgery or radiotherapy. Many surgical oncologists resist participation in such trials, and trials that have been done have come long after a procedure is introduced. The history of surgical techniques used in treating breast cancer illustrates this. The proposal that a lesser operation be used in place of a radical (Halsted) mastectomy was published in 1948. Not until 1967 was a trial carried out. Even today, though the practice has gradually declined, many women undergo radical mastectomy when a modified procedure would be equally effective and less disfiguring (see the section "Breast Cancer" above and ref. 226).

The literature on the impact of RCTs in surgery is limited, considering the size of the field. One volume, *Costs, Risks, and Benefits Surgery,* covers a wide range of topics in surgical innovation and evaluation, including RCTs (28). The editors conclude with a series of recommendations, including those for improving the study of surgical procedures (see ch. 6).

Bunker and colleagues (29) studied the introduction and evaluation of four modern surgical procedures, three that were eventually assessed by RCTs. They note the particular problem of carrying out RCTs of new therapies for conditions that previously had no effective therapy of any kind. Withholding treatment in these cases can pose difficult ethical questions. The use of shunt surgery for portal hypertension is one example. After decades of use, the procedure was subjected to evaluation by RCT only because of two developments: the recognition that the surgery had a serious side effect (encephalopathy), and the extension of the use of the operation beyond its original indications. The uncertainty about the use of the surgery for new indications, using it prophylactically rather than just therapeutically, led to RCTs with the newly indicated group of patients. After these trials showed shunt surgery to be ineffective prophylactically, further trials demonstrated its lack of efficacy for its original therapeutic uses.

Three case studies in *Assessing the Efficacy and Safety of Medical Technologies* discuss surgical procedures that require evaluation, largely because RCTs of them have been inadequate or simply not done (225). These three case studies are summarized below.

Tonsillectomy, the third most common surgical procedure in U.S. hospitals, is thought by many physicians to be overused. Reports of tonsillectomy reach back as far as 600 B.C., yet the first RCT of the procedure in this country began in 1973. Tonsillectomy differs from some other procedures with long histories, such as cast application for bone fractures, in that its efficacy is not obvious and the indications for use not well understood. The National Institutes of Health (NIH) sponsored a workshop in 1973 on Tonsillectomy and Adenoidectomy that recommended a nationwide multicenter RCT. That idea was later endorsed by another NIH-convened group, the NIH Ad Hoc Advisory Panel on Tonsillectomy and Adenoidectomy. In 1978, a third group did not agree to go ahead with the trial.

Appendectomy is another frequently performed surgical procedure that has not been evaluated by an RCT in this country. The different rates of appendectomy in different regions of the country (from 100 to 620 per 100,000 for 1965-73) and evidence from other parts of the world provide strong support for the need to understand the appropriate use of this procedure. The OTA report concluded that an RCT might be warranted in view of "strong evidence suggesting that appendicitis may be treated with substantially fewer appendectomies without increased loss of life."

Hysterectomies are performed for a wide variety of conditions, including the traditional indications of premalignant states, localized cancers, descent and prolapse of the uterus, and obstetric catastrophes (e.g., functional problems). Performed in over 600 per 100,000 women each year, this major operation is more frequently performed than any other. In assessing the costs, risks, and benefits of elective hysterectomy, Korenbrot and colleagues reviewed studies indicating that at least 30 percent of hysterectomies performed were not justified by medical indications alone (125). The implication, though unprovable, is that most were performed for sterilization or cancer prophylaxis. Lack of clarity about the procedure's appropriate indications and the substantial risks and poorly known aftereffects of the surgery itself emphasize the need for controlled trials. In 1978, OTA was unable to identify any clinical trial of hysterectomy in this country.

Neurosurgery

Haines has recently examined RCTs in neurosurgery based on an exhaustive search of the English language literature (103). In an earlier paper, he reviewed 4,685 scientific articles appearing between 1944 and 1977 in the *Journal of Neurosurgery,* finding that only 18 could be classified as controlled clinical trials, and of those, 10 used random allocation procedures (104). One of the ten used blinding procedures. His later, more extensive review (103) identified a total of 51 RCTs of neurosurgical procedures, adjuncts to neurosurgical procedures or medical treatment of neurosurgical diseases. Half these studies were published after 1977. Most of the studies (61 percent) were of adjuncts to surgical therapy (e.g., radiation and chemotherapy for malignant primary brain tumors), 15 directly tested a neurosurgical procedure, and 5 nonsurgical therapy, such as antibiotic treatment of shunt infection.

The increased use of RCTs in neurosurgery is encouraging, but Haines asks: "Have any important questions been resolved by such studies?" He answers with a qualified "no." A large percentage of the trials were methodologically inadequate and permitted no conclusions. The well-conducted studies, however, though they failed to put important questions to rest, did gather important information about the natural history of diseases. The case has been made for more definitive trials in this field, some of which are under way. In neurosurgery, and probably in other surgical areas, the quality of trials is a serious problem. Statisticians have not been routinely involved in design, which proves to be a major determinant of trial quality (105). Progress has been relatively slow, and will come only with surgeons' greater appreciation of the value of RCTs.

Haines reports a case of negative results in small RCTs with low statistical power, that encouraged an unwarranted decline in a neurosurgical practice (105). A standard practice in the late 1970's was the use of antifibrinolytic agents in treating patients with subarachnoid hemorrhage from ruptured intracranial aneurysm. The purpose of the treatment is to prevent recurrent hemorrhage during the waiting period between first hemorrhage and surgery. Haines reports that three recent reviewers have seriously questioned the efficacy of this therapy, based on the evidence from RCTs, and have suggested that antifibrinolytic agents may aggravate another problem, vasospasm. Haines' reassessment of the RCTs yields a different conclusion. The four trials that showed the treatment was ineffective all had a less than one chance in three of finding a 50 percent better outcome in the treated group, if such a difference existed. The three studies with the greatest statistical power showed some benefit from the therapy, and little evidence for its aggravation of vasospasm. Haines concludes that discarding antifibrinolytic therapy is premature. He recommends further clinical trials to study both its efficacy and safety, in studies that are well designed and large enough to produce significant answers.

RCTs IN OTHER FIELDS

Chalmers and colleagues (40) have been engaged over about the last 5 years in the development of a computerized data base of RCTs. As of 1982, about 2,700 RCTs were entered, indexed by groupings of the International Classification of Diseases (WHO, 1977). From their data base, Chalmers and colleagues have identified common disease states for which a relatively large number of RCTs are available, and have evaluated the quality of the trials according to an index they have developed (see ch. 4, "Quality of RCTs"). Where possible, they have synthesized the results of studies to draw conclusions about therapies tested. Topics addressed have been: surgical therapy of duodenal ulcer, early mobilization and discharge of acute myocardial infarction patients, antithrombotic agents in acute myocardial infarction, cost and efficacy of the substitution of ambulatory for inpatient care, treatment of acute alcohol withdrawal, treatment of acute infections and alcoholic hepatitis, nephrology, tropical diseases, effects of steroids in the gastrointestinal tract, and emergency diagnosis and treatment of gastrointestinal hemorrhage.

The degree to which RCTs are used in different fields of medicine varies greatly, hence the impact of RCTs must vary. Certain areas have not been mentioned specifically in this chapter, for instance pediatrics, and obstetrics and gynecology. In these areas too few RCTs have been conducted to allow much impact. While it is easy to focus on deficiencies of studies that are done, it is more important though more difficult to identify medical fields which lack RCTs altogether. Very little has appeared in the literature in this regard, except in the case of surgery, which was reviewed in this chapter.

6.

Improving the Impact of Randomized Clinical Trials

6.

Improving the Impact of Randomized Clinical Trials

Throughout the course of this background paper, opportunities have been identified to improve the impact of randomized clinical trials (RCTs) on medical practice and for the expanded use of RCTs in policymaking. Potential improvements fall in the following categories: 1) the quality of RCTs that are carried out, 2) the dissemination of information from RCTs, 3) the overall system of assessing medical technologies, 4) the use of RCTs for policy decisions, and 5) the use of RCTs in specific medical fields. The following suggestions have appeared in the published literature or arose in discussions with individuals during the course of preparing this paper.

IMPROVING THE QUALITY OF RCTs

If RCTs are to have more influence on health policymaking and medical practice, the way they are conducted needs to be improved in several ways: they should adhere to known principles of design, including statistical and other methods; they should be further improved through greater support for research in RCT methods; journal editors should impose stricter standards for RCT reports; and they should increasingly take the form of multicenter RCTs.

The Broader Application of Good Experimental Methods

Basic principles on which good RCTs depend are known. They are not always applied, however. To the extent that lack of application is a consequence of lack of knowledge of good methodology, improvements can be made at various points in the medical education system: in medical school education; in residency programs; and in continuing medical education. Outside of medical education, funding agencies, notably the National Institutes of Health (NIH), could be more assiduous in requiring good study design for funding approval, and even in providing assistance to improve deficient study designs that are submitted.

There has been some movement toward teaching quantitative methods in medical schools, but progress is slow. A suggestion for speeding up the process is to involve the American Association of Medical Colleges (AAMC) in developing curricula for teaching research methods, including RCTs.

The requirement for new drug approval gives the Food and Drug Administration (FDA) considerable potential leverage over the conduct of RCTs. This leverage could be used to improve the adequacy of RCTs on medical devices as well as drugs. FDA has developed, in addition to regulations, a series of guidelines for the conduct of RCTs. Adherence to these guidelines implies that results of the study will be considered as part of a New Drug Application. FDA's guidelines are quite general and set only minimal methodological standards. The guidelines could be strengthened to include standards for designing, implementing, and reporting trials. Standards for sample size, length of followup, and completeness of followup, might be considered as well as reporting requirements. Drug companies and medical device manufacturers and the groups with whom they contract to conduct RCTs are likely to be very responsive to FDA guidelines (189).

In part, a lack of faculty qualified in quantitative methods may hamper the teaching of these methods in medical schools. NIH has a program of career development awards in medicine, but none in the field of biostatistical methods. Making such awards might further the teaching of quantitative methods in medical schools (255).

Providing assistance for designing sound RCTs by granting agencies is not a new idea. The National Eye Institute (NEI), in the early years of encouraging RCTs, made small planning awards to those with good ideas, but in need of statistical and methodological assistance for RCTs. Such a program could be targeted to areas of medicine in which RCTs still are not widely used.

Improving Statistical Methods Through Research

The application of known statistical principles in trials would go a long way toward improving them. There is also scope for improving the methods themselves. The RCT is a relatively recently developed method, and deserves to be developed to the fullest.

The Federal research establishment does not now systematically support research to develop biostatistical methods. NIH has no study section to review grant applications in biostatistics and clinical trial methodology, and therefore relies on ad hoc groups. As a result, these groups may not be made up of those most qualified to review the grants received. A permanent study section would likely be more carefully chosen, and its existence might encourage more grant proposals to develop innovative methods in clinical research. Further, improving RCTs will depend on advances in biostatistical methods.

Applying Stricter Editorial Standards

Because publication is a critical part of the RCT process, and publications are important to the careers of researchers, journal editors wield a powerful tool in their standards for acceptance. Many have argued that these standards should be more rigorous. Curtis Meinert, the editor of *Controlled Clinical Trials,* proposes that the following information should be required in a report for publication (159):

- the source of funding for the trial and an indication of whether the reported results are a subgroup of a larger data set;
- a list of the treatment groups and the rationale for the choice of treatments;
- a description of the method to allocate patients to treatment groups, including reference to the blinding used in each group (i.e., none, single or double blinded);
- the safeguards used in the trial to protect patients informed consent and privacy;
- the criteria used to exclude patients from the trial;
- the criteria used to include patients in the trial;
- the rationale for the number of patients studied, including a statement of assumptions used in calculating the sample size;
- a statement of the length of time required to complete patient enrollment;
- a description of the population from which patients were selected;
- a description of the baseline and followup examination schedule;
- a specification of the key outcome variable(s);
- the descriptive information on the baseline comparability of the treatment groups;
- the number of patients assigned to each treatment group;
- the level of patient compliance achieved in each treatment group;
- the number of patients followed to the end of the study or to death;
- the number of deceased patients;
- the number of patients unable or unwilling to return for followup examinations, including a count of the number who could not be located at the end of the study;
- a description of quality control procedures used in collecting data;
- a description of the methods of analysis, including an indication whether the reported p values resulted from a single or repeated evaluation of the data; and
- a discussion of the power of the study.

Encouraging Multicenter RCTs

Multicenter RCTs should be encouraged in situations where increased sample size and a more heterogeneous population are assets. Strategies to overcome some of the difficulties of multicenter trials should be developed.

Carrying out multicenter trials requires that a large number of investigators cooperate, however, and the present incentives for individuals to do so are low, regardless of their interest in the study. Reports of multicenter RCTs often cite the author as the cooperative group or may list a dozen names, sometimes at the report's end. Such forms of citation do little for the professional standing of researchers in academic settings. Meinert (159) suggests the following to overcome some of these disincentives:

1. Encourage investigators to participate by recognizing participation in promotion criteria for academic faculty.
2. Allow greater flexibility for participating investigators to carry out related investigations which they can publish under their own names.
3. Award greater recognition to the field of clinical trials as a professional activity and not just as an adjunct to treating patients.

Noting the contributions of community hospital physicians in recent trials (ch. 4, "Multicenter Trials," and ch. 5, "Impact of the Cooperative Oncology Groups"), Cease (38) argues that such participation is, in fact, continuing medical education (CME). He suggests that CME credit be awarded for a certain level of participation, to serve both as recognition of achievement and as an incentive to participate.

IMPROVING THE DISSEMINATION AND USE OF RCT RESULTS

The results of RCTs can be useful in several ways. One well-designed, well-conducted RCT can provide convincing evidence for a change in practice. In that case, the results should be known to clinicians so that they may change their behavior accordingly. The results of another RCT might not be so unequivocal. They might not be the basis for altering practice immediately. If there are enough other trials on the same subject, the results taken together might suggest a clearer answer. That situation calls for some type of synthesis, perhaps a meta-analysis of RCTs. Publication of the synthesis results might then be the basis for changing clinical practice.

An RCT may confirm that current practice is indeed effective, or more effective than a newer practice, and those results should be known to physicians in the appropriate fields.

In addition to providing guidance for medical practice, RCTs may contribute to further research, either in the design of future RCTs or in other types of research. In that case, it is researchers who will benefit from knowing the results of the RCT.

Finally, information about patient treatment techniques, other than the final result, is generated in RCTs.

Optimal strategies for disseminating information from RCTs will differ depending on which group needs to know about the results, and what aspects of the results are most relevant. Two basic approaches are needed:

1. an active dissemination effort, trying to reach those who need to know with the results, and
2. facilitating access to RCT results for those who want to find out.

The traditional and still most important method of disseminating research results of any kind, including those of RCTs, is through publication in technical journals. This may be sufficient for trials that are not of great clinical significance. For those which clearly point the way for changing medical practice, however, a single publication, even in the most prestigious medical journals, may not reach those who need to know, namely the practitioners in the field of the trial or general practitioners who sometimes or frequently work in the areas. In some cases, interesting results in treating diseases of high public visibility may lead to publicity in the mass media, but such occurrences are rather rare. Medical news publications report on a greater proportion of research results of clinical interest. Beyond those routes, there must be greater initiative on the part of investigators and perhaps funding agencies to disseminate findings from RCTs.

Pharmaceutical companies make the most direct use of RCT results in advertising their products. Implicit in their statements about safety and efficacy is the backing of RCT results. They advertise both in widely read subscription journals and in widely distributed "throwaway" publications. In addition, their representatives personally visit physicians and institutions. Together these public relations achieve widespread awareness of a company's products.

FDA might also draw clinicians' attention to RCT results if they more formally included RCT results in the Physicians' Desk Reference (PDR) drug inserts. Inclusion of a brief account of supporting RCTs, indicating the methods, results, and limitations of the trials would provide clinicians with a basis for their own critical analysis before prescribing a drug (189).

Government and private funding agencies probably cannot match the efforts of pharmaceutical companies, and to do so might not be desirable. Nevertheless, they could greatly improve in this regard. The National Heart, Lung, and Blood Institute (NHLBI) leads such funding bodies in disseminating results (ch. 5). Its use of the medical news media, workshops, and meetings could serve as a model for other organizations.

The medical specialty societies also help disseminate information. Most active at present is the American College of Physicians (the association of physicians who have demonstrated competence in internal medicine). These societies should be encouraged to educate members both about RCT methods and about the results of specific RCTs.

The institutes of NIH, to varying degrees, also disseminate information by holding meetings at the NIH campus, sponsoring sessions at meetings of specialty societies, and sponsoring and disseminating the results of consensus development conferences.

The National Cancer Institute (NCI) has begun a program to facilitate access to active trials in clinical cancer research. The "PDQ" system is an international computerized data base accessible to patients and physicians, containing protocols of clinical research (see ch. 5).

Chalmers and his colleagues have begun a major effort to facilitate access to RCT results in all fields of medicine. Having collected published reports of RCTs for a number of years, as of 1982 a total of nearly 3,000, they have begun computerizing this information so that investigators and clinical physicians can have ready access to data on RCTs in specific areas. This is not possible through any existing data base. Included in each entry is an evaluation of the trial by Chalmers' quality index (ch. 4). The system will facilitate the synthesis of results from trials in many fields.

With the proliferation of personal computers, data bases such as Chalmers has established and NCI's PDQ system should be available to practicing physicians. Funding agencies and the preparers of data bases could profitably undertake efforts to ensure that clinically relevant research, including RCTs is readily accessible to clinicians with personal computers.

Probstfield and his colleagues (185) have identified a failing in dissemination of information from RCTs which has rarely been addressed. It is that "the methodological knowledge gained from clinical trials cannot at present be systematically transferred to clinical practice." The areas that Probstfield and his colleagues have identified in which clinical trial methods can contribute to clinical medicine are: clinic operations and management, the quality control of clinical practice, patient adherence to therapeutic regimens, and staff education. Information about these subjects may be available even before the trial is over. The authors suggest some steps that would improve the access to and use of information from clinical trials:

- a computerized retrieval system at some central source for clinical trials methods must be developed, maintained and consistently updated with appropriate cataloging of new developments;
- scientists in clinical trials must make additional efforts to recognize and to highlight in specific publications the methodology which is relevant for clinical practitioners;
- a systematic transfer of the clinical trials methodology literature to that literature read by the clinical practitioner is crucial. This transfer may require brief summaries of methods published regularly in journals with appropriate circulation and readership; and
- facilities on a national or regional basis must be developed to train clinical practitioners in methods validated in clinical trials.

IMPROVING THE ASSESSMENT OF MEDICAL TECHNOLOGIES

The results of RCTs should have the greatest impact possible. This entails developing a rational means to set priorities in funding research given the limited dollars available. The priority criteria should take into account which technologies are important for health policymaking and medical practice.

NHLBI's decisionmaking procedure for large-scale RCTs is one model for a mechanism to set priorities (see ch. 5, "NHLBI and RCTs"). Bunker and Fowles' (27) "Institute for Health Care Evaluation" (IHCE) proposes another model for this mechanism to improve the evaluation of medical technologies in all its phases (see ch. 3). One important function of IHCE would be to set research priorities.

Perry (178) proposes that a "Center for Assessment of Health Care Technology" be established in the private sector. Like IHCE, this Center would be a nonprofit organization funded by several sources: "private foundations, private third-party payers and health insurance alliances, group health and hospital associations, and corporations and labor unions with major health insurance programs for employees." Perry adds, "it is also conceivable that funds could be obtained under contract from HCFA [the Health Care Financing Administration] for evaluations to be used in coverage decisions and from other Federal or State agencies requiring similar services." Though Perry applauds related activities in the private sector such as the Clinical Efficacy Assessment Project of the American College of Physicians (ch. 3) and other efforts sponsored by the medical community, he thinks they cannot replace the impartial assessment that is possible by an organization without special interests—e.g., the proposed center.

Suggestions have been made to increase the efficiency of the process leading up to clinical trials. This would require the earlier identification of technologies that will need assessment and the improved use of information gathered prior to any RCT. If a new procedure is first tried on patients at various locations around the country, for instance, the data collected on each case could profitably be standardized and pooled, and perhaps placed in a data bank. None of these procedures are generally followed today, and many more patients than those required may undergo the procedure before one center or group has sufficient data to plan a good trial.

Mosteller and Weinstein (164) have proposed a method to evaluate the costs, risks, and benefits of clinical trials before they are carried out. Their technique is proposed to improve the rationality of spending for medical research and evaluation. In essence, the evaluation attempts to predict what the impact of doing a trial may be and with that information to decide whether the trial would be worthwhile. The authors lay out a large number of assumptions and uncertainties in formulating their model. One of its valuable aspects is that it forces a wide range of probable impacts to be considered, not only the potential benefits and risks of the procedure, but also the potential value of new knowledge gained about the disease, clinical trial methods, and health services delivery, for example. Such issues as possible misapplication of the procedure, the probability of widespread diffusion of a technology before the study is completed, and other relative unknowns figure in the evaluation.

An additional benefit of the evaluation is that it facilitates actual assessment of impact after a trial is finished, a task which has seldom if ever been accomplished with total success.

THE USE OF RCTs IN POLICY DECISIONS

Some have suggested that the trend of using RCT results in making policy should be encouraged. In large part they refer to decisions about coverage of medical services by third-party payers, both public and private. RCT results might be more useful for policy decisions if there were greater interaction between third-party payers and funding agencies. This would help to focus RCTs on health issues directly relevant for policy, and more generally, to make all RCTs more relevant to policy. The latter could be accomplished by including components on cost, for instance. Contributions to funding RCTs, discussed in the section below, might help in this effort.

At lower levels of policymaking, RCT results could be used more extensively by hospitals and other medical institutions in decisionmaking about their services.

Funding of RCTs

NIH spends more money funding clinical trials than any other institution in the United States, and perhaps, in the world. In the last year for which figures are available, NIH spent 4.3 percent of its total budget on clinical trials (not all are RCTs; see ch. 2). In 1975, it spent 5 percent of the total budget on clinical trials. The trend since 1979 is unknown, though there is reason to believe the share spent on clinical trials has diminished (78). Even at the 1979 level, "136 million of an approximately $3 billion total budget for NIH, shows a rather small commitment to testing the results of years of basic and applied research" (17).

Apart from increasing NIH funds for clinical trials, funding can be increased to the extent the costs of RCTs can be distributed more fully within the health care system. Third-party payers currently reimburse for some costs of patient care and hospitalization in RCTs. That share could be increased (see ch. 2 for a full discussion of RCT funding by third-party payers). Some progress has been made, and efforts are under way to facilitate greater participation in RCT funding by health insurers.

For the first time, as a result of the 1983 Social Security Act Amendments, HCFA will be allowed to fund RCTs. Presumably they will use that capability to answer questions of direct policy relevance to the Medicare and Medicaid programs. Not only does HCFA have the opportunity to provide useful information, but their activities, if successful, may stimulate similar commitments among private third-party payers.

IMPROVED USE OF RCTs IN SPECIFIC FIELDS

Suggestions have been made to extend and improve the use of RCTs in specific areas of medicine and for specific types of technologies. These are discussed below.

Surgery

The uses and limitations of RCTs in surgery are discussed in chapter 5. The recommendations made by Bunker and colleagues in *Costs, Risks, and Benefits of Surgery* (28) are reproduced here:

> **Recommendation I**
>
> Appropriate studies of the effectiveness of surgical treatment should be carried out for selected conditions, particularly those where uncertainty leads to professional disagreement.
>
> . . . *Improving techniques for evaluation.* At the same time that studies using currently available methods must go forward, we have seen the need to improve our ability to conduct these urgently needed studies. A major problem is our presently inadequate information system. Separate records are kept for each patient by each physician or institution caring for him. In 1977 it is possible to identify outcome as related to an operation or other treatment only if the treatment and the observed outcome occur during a single continuous hospitalization. Even under these circumstances the standard medical record is not designed for easy information retrieval or the pooling of information across patients to study populations. It is frequently nearly impossible to document the treatment and health status found at previous examinations, especially if a different hospital or physician were responsible. Existing data cannot determine long-term outcomes or the end-result of surgery. Thus we are unable to find out, except for selected conditions such as malignant tumors and end-stage renal disease, how many patients survive one or more years after a particular operation. We cannot determine how many patients have been relieved of the condition leading to the operation, or how many fully

recovered from the effects of anesthesia and surgery and been able to return to full, pre-illness activity.

We are now able to perform useful cost-risk-benefit analyses, but present techniques need to be improved; for example, we are probably not sufficiently aware of second order effects or unanticipated consequences of proposed new policies. Perhaps we can learn to anticipate such "unanticipated consequences." Careful work still remains to be done on methodology of experimental design. It is not sufficiently widely recognized how long it takes to design an informative clinical trial or how difficult it is to execute the design once it has been chosen. We do not yet know enough about randomized trials and their consequences, their weaknesses, strengths, and costs compared with their alternatives. We still are not sure enough of when we should trust an observational study. We do not know how to combine epidemiology and observational and experimental information. We have not dealt with the ethical issues surrounding human experimentation and are still shouting at one another from fixed positions. We have not reviewed the complexities of our ethical problems in enough detail or sophistication.

Recommendation II

Our grasp of the components of cost-benefit analysis and their interrelations, the values of the various data gathering techniques, and our understanding of the ethics of data gathering must be improved by theoretical and empirical work and by continued discussions in the public forums.

. . . *Improving medical capabilities for evaluation.* In addition to assessing the efficacy of many existing treatments, we need to develop a policy for the introduction of new medical and surgical technology. Thus among the studies encouraged in Recommendation II, we would include further historical studies of past successes and failures. We call particular attention to two recently published studies. One, the "Study on Surgical Services for the United States"(172), includes a survey of the major surgical advances of the past quarter-century and the research on which these advances were based. The second, entitled "Scientific Basis for the Support of Biomedical Science" (54), examines in detail the research basis for recent advances in the surgical and medical treatment of cardiovascular and pulmonary diseases. Studying only successes or failures can have weaknesses that a balanced approach may avoid.

Even when the technology and data may be available, the current methods need to be more widely understood in the medical research and medical policy communities as well as among medical students and their teachers. Naturally, we cannot expect all to be experts. But physicians themselves must be better educated in the analytic techniques necessary for them to make a more informed discrimination among therapeutic programs or techniques, and they must be educated in the economic, social, and epidemiological principles of medical care which will allow them to participate as leaders of society in advising on or helping to make priority decisions.

Recommendation III

These principles of cost-benefit evaluation should be included as an integral part of the medical school curriculum; and their application to the assessment of the efficacy of medical care should be incorporated into clinical practice and continuing medical education.

We note in particular that medical students at the beginning of their clinical training may feel little pressure to know much about the design of clinical trials or of policy analysis. Later, when working in the hospital and trying to read and appraise results presented in research papers or in participating in research, knowledge of these matters absorb the young physician's attention. Thus, we stress continuing education.

Improving public understanding. In addition to educating itself, the medical community has an obligation to inform the public. Here we would note a distinction made by the sociologist Paul Lazarsfeld between advising and deciding. After data are gathered by good methods and carefully analyzed, the scientist or physician needs to advise the client, here the community, about the findings. The community takes this advice and tempers it with political, legal, social, and moral considerations and then decides. We should improve our advice so that it will be useful in the decision process.

Recommendation IV

Information on outcomes as well as costs of medical care should be routinely formulated in a manner suitable for presentation to the public.

Cancer Research

The use of RCTs in cancer research could be improved through better statistical analysis of the potential value of a trial, and through directing them more frequently to research in cancer prevention.

Zelen, Gehan, and Glidewell (258) suggest that the following conditions be met for a trial to be done:

1. Do not initiate a definitive clinical trial unless there is a reasonable a priori probability greater than 0.05 that a clinically important gain may exist. One way of interpreting this rule or behavior is to carry out pilot studies before launching a definitive study. If the pilot studies are encouraging, then proceed with a large comparative study.
2. Comparative trials should be planned with a minimum of 100 to 200 patients per treatment. Trials with fewer patients are likely to produce more false positive results than true positive results.
3. All positive results should be independently confirmed. This will lower the false positive rate and raise the true positive rate. Physicians in practice should exercise caution in adopting a new therapy if there is no independent confirmation.

Greater emphasis on cancer prevention is warranted in RCTs. The first major trial in primary prevention is now under way. Sponsored by NCI, it is testing beta carotene, a precursor of vitamin A, as a cancer inhibitor. One important cancer screening technique, the use of mammography to detect breast cancer, has been carefully evaluated in RCTs. Several trials of lung cancer screening are now ongoing. The survival rate of those with the most common types of cancer—lung, gastrointestinal, and breast cancers—has not improved greatly since the 1950's (226). Thus, the detection and treatment of cancer at its early stages seems a reasonable immediate goal. Though admittedly expensive and administratively complex, the larger trials necessary to evaluate screening procedures would be worthwhile. They might compare favorably in the information they produce with large-scale secondary prevention trials in cardiovascular disease.

CONCLUSION

This paper has reviewed the available literature about the impacts of RCTs. The use of RCTs themselves is a relatively recent development, beginning only in the middle of this century and still gaining in popularity. Concern about the impacts of RCTs has come even more recently, and ideas for improving or increasing these impacts have been little voiced. Based on the small literature now available, additional effort could be profitably directed toward understanding the impacts of RCTs, and devising methods for maximizing their usefulness in health policymaking and in influencing medical practice. RCTs could play a greater role in the national use of medical technology at all levels of decisionmaking.

Appendixes

Appendix A.—Acronyms and Glossary

Acronyms

ADAMHA — Alcohol, Drug Abuse, and Mental Health Administration (PHS)
ALL — acute lymphoblastic leukemia
AMIS — Aspirin Myocardial Infarction Study
AML — acute myelocytic leukemia
BP — blood pressure
CABG — coronary artery bypass graft (surgery)
CAT — computed axial tomography
CCU — coronary care unit
CDP — Coronary Drug Project
CEA — cost-effectiveness analysis
CME — continuing medical education
CSP — Cooperative Studies Program (VA)
CT — computed tomography
DBP — diastolic blood pressure
DES — Drug Efficacy Study
DES — diethylstilbestrol
DHHS — Department of Health and Human Services
DOD — Department of Defense
DRS — Diabetic Retinopathy Study
ECOG — Eastern Cooperative Oncology Group
FDA — Food and Drug Administration
HBGM — home blood glucose monitoring
HCFA — Health Care Financing Administration (DHHS)
HCT — historical control trial
HMO(s) — health maintenance organization
ICU — intensive care unit
IHCE — Institute for Health Care Evaluation
MRFIT — Multiple Risk Factor Intervention Trial
NAS — National Academy of Sciences
NCHCT — National Center for Health Care Technology (PHS)
NCI — National Cancer Institute (NIH)
NEI — National Eye Institute (NIH)
NHLBI — National Heart, Lung, and Blood Institute (NIH)
NIAAA — National Institute on Alcohol Abuse and Alcoholism (ADAMHA)
NIADDK — National Institute of Arthritis, Diabetes, and Digestive and Kidney Diseases (NIH)
NIAID — National Institute of Allergy and Infectious Diseases (NIH)
NIDA — National Institute on Drug Abuse (ADAMHA)
NIEHS — National Institute of Environmental Health Sciences (NIH)
NIGMS — National Institute of General Medical Sciences (NIH)
NIH — National Institutes of Health
NIMH — National Institute of Mental Health (ADAMHA)
NINCDS — National Institute of Neurological and Communicative Disorders and Stroke (NIH)
NRC — National Research Council (NAS)
NSABP — National Surgical Adjuvant Project for Breast and Bowel Cancers
OHTA — Office of Health Technology Assessment (PHS)
OMAR — Office for Medical Applications of Research (NIH)
OTA — Office of Technology Assessment (U.S. Congress)
PHS — Public Health Service (DHHS)
POSCH — Program on the Surgical Control of the Hyperlipidemias
RCT(s) — randomized clinical trials
TAR — Treatment Assessment Research
VA — Veterans Administration

Glossary

Apheresis: A procedure that separates the blood into its basic components (red cells, white cells, platelets, and plasma) and selectively removes one or more of these components from the blood for the purpose of curing, alleviating, or treating a disease and its symptoms.

Blinding: Keeping secret which treatment is assigned to participants in randomized clinical trials. When only the patient is kept unaware of his or her treatment assignment, the study is "single-blind;" when the person administering treatment (e.g., the physician) also is unaware, the study is "double-blind." Additional layers of blinding can be added—e.g., when a third individual, the evaluator of outcome, also is unaware of treatment assignments.

Chemotherapy: The treatment of disease by chemical agents.

Concurrent controls: In a clinical trial, individuals given a "control treatment" during the same time period as experimentally treated individuals, usually used to refer to individuals not formally enrolled in the trial.

Consensus: General agreement on a subject, not necessarily grounded in fact.

Control group: In a randomized clinical trial, the group receiving treatment with which the group receiving experimental treatment is compared. The control treatment is generally a standard treatment, a placebo, or no treatment.

Crossover: In a randomized clinical trial, switching of treatment during the course of the trial. Crossovers can be planned as part of the trial method, or unplanned, a consequence of an individual's changing medical condition.

Device (medical): Any physical item, excluding drugs, used in medical care (including instruments, apparatus, machines, implants, and reagents).

Disease prevention: The aversion of disease, traditionally characterized as primary, secondary, and tertiary prevention. Primary prevention aims at avoiding disease altogether. Secondary prevention strategies detect disease in its early stages of development, with the hope of improving outcome. Tertiary prevention attempts to arrest further deterioration in individuals who already suffer from a disease.

Drug: Any chemical or biological substance that may be applied to, ingested by, or injected in order to prevent, treat, or diagnose disease or other medical conditions.

Effectiveness: Same as efficacy (see below) except that it refers to average or usual conditions of use.

Efficacy: The probability of benefit to individuals in a defined population from a medical technology applied for a given medical problem under ideal conditions of use.

Experimental group: In a randomized clinical trial, the group receiving the treatment being evaluated for safety and efficacy. The experimental treatment may be a new technology, an existing technology applied to a new problem, or an accepted treatment about whose safety or efficacy there is doubt.

External controls: In a clinical trial, individuals given a "control treatment" with which the experimentally treated group is compared, but who are not formally enrolled in the trial. External controls may be historical or concurrent.

Historical controls: In nonrandomized clinical trials, individuals treated with a "control treatment" outside the study proper, at some time previous to the trial, against which the experimentally treated individuals are compared.

Mammography: X-ray examination of the breast, used as both a screening procedure on apparently healthy females, and as a diagnostic procedure in clinical situations to detect breast cancer.

Medical technologies: Drugs, devices, and medical and surgical procedures. The organizational and supportive systems through which medical care is provided are part of medical technology in its broadest sense, but are not discussed in this report.

Minimization: In randomized clinical trials, a method of patient allocation which seeks to minimize different distributions of prognostic factors between treatment groups without creating mutually exclusive subgroups.

p value: In a randomized clinical trial, the probability of concluding that there is a difference between the treatment groups when, in fact, there is none. Also called "Type I error" or "alpha" and commonly called the "level of statistical significance," analogous to "false positive."

Phase I, II, and III drug trials: The sequence of studies in human beings required for new drug approval by the Food and Drug Administration. Phase I includes studies in a small number of relatively healthy patients or normal volunteers to determine safety and pharmacologic effects. Phase II includes controlled clinical trials to determine appropriate doses, safety, and effectiveness in a total of about 200 patients. Phase III trials are usually randomized clinical trials (RCTs).

Placebo: A drug or procedure with no intrinsic therapeutic value which mimics the drug or procedure being tested in a randomized clinical trial. A placebo is used in control groups as a means to blind patients and investigators as to whether an individual is receiving the experimental or control treatment.

Prognostic factors: Symptoms, signs, or characteristics of an individual that are known to be predictive for certain disease outcomes.

Random allocation: In a randomized clinical trial, allocation of individuals to treatment groups such that each individual has an equal probability of being assigned to any group.

Randomized clinical trial (RCT): An experiment designed to test the safety and efficacy of a medical technology in which people are randomly allocated to experimental or control groups, and outcomes compared.

Risk: A measure of the probability of untoward outcomes occurring, and the severity of the resultant harm to health of individuals in a defined population associated with use of a medical technology, applied for a given medical problem under specified conditions of use.

Safety: A judgment of the acceptability of risk in a specified situation.

Statistical power: In a randomized clinical trial, the probability of detecting a difference between the treatment groups when one does exist. Failure to detect an effect is called "Type II error" or "beta," analogous to "false negative."

Statistical significance: See p value.

Stratification: In randomized clinical trials, the categorization of individuals for the purpose of adjusting the groups to take into account unequal distribution of characteristics of prognostic importance. Stratification may be used during patient allocation, creating subgroups within which individuals are randomized to treatments; or stratification may be applied during data analysis to statistically adjust for differences between the groups.

Synthesis: The integration of findings from different studies and the development of generalizations based on their results.

Type I error: See p value.

Type II error: See statistical power.

Validity: A measure of the extent to which an observed situation reflects the "true" situation. *Internal validity* is a measure of the extent to which study results reflect the true relationship of a technology to the outcome of interest in the study subjects. *External validity* is a measure of the extent to which study results can be generalized to the population which is *represented* by individuals in the study.

Appendix B.— Acknowledgments

The people listed below contributed in some way to this study. Some provided information during the course of the study and others made suggestions for improvements during the review phase.

John Adams
  Pharmaceutical Manufacturers Association
John C. Bailar III, M.D.
  Public Health Service
John Ball, M.D.
  American College of Physicians
Bernard S. Bloom, Ph.D.
  University of Pennsylvania
Queta Bond, M.D.
  Institute of Medicine
Richard J. Brand
  University of California at Berkeley
  School of Public Health
Paul Bryan
  Food and Drug Administration
John P. Bunker, M.D.
  Stanford University School of Medicine
David Byar
  National Cancer Institute
Col. Craig Canfield
  Walter Reed Army Medical Center
Paul L. Canner, Ph.D.
  Maryland Medical Research Institute
Bruce A. Chabner, M.D.
  National Cancer Institute
The Commonwealth Fund
  New York
Dennis Cotter
  Office of Health Technology Assessment
Robert Edelman, M.D.
  National Institute of Allergy and Infectious Diseases
Fred Ederer
  National Eye Institute
Bernard Fisher, M.D.
  University of Pittsburgh School of Medicine
Joseph Fleischer
  Maryland Medical Research Institute
Peter Fox
  Lewin & Associates
Edward D. Freis, M.D.
  Veterans Administration
Lawrence M. Friedman, M.D.
  National Heart, Lung, and Blood Institute
Robert Gordon, M.D.
  National Institutes of Health
Richard Greene
  Veterans Administration
Stephen J. Haines, M.D.
  University of Minnesota Medical School
John James
  National Institutes of Health
Lisa Kaufman, M.D.
  American College of Physicians
Christian R. Klimt, M.D., Dr.P.H.
  Maryland Medical Research Institute
Genell Knatterud, Ph.D.
  Maryland Medical Research Institute
Richard Kopanda
  Alcohol, Drug Abuse, and Mental Health Administration
Phil Lee, M.D.
  University of California at San Francisco
  Health Policy Program
Alan Lichter
  National Cancer Institute
Harry M. Marks
  Harvard Medical School
Alan E. Mayers, Ph.D.
  National Center for Health Services Research
Michael P. McQuillen, M.D.
  Medical College of Wisconsin
Curtis T. Meinert, Ph.D.
  The Johns Hopkins University School of Hygiene and Public Health
Valerie Miké, M.D.
  Memorial Sloan-Kettering Cancer Center
Lincoln E. Moses
  Stanford University School of Medicine
Duncan Neuhauser, Ph.D.
  Case Western Reserve University School of Medicine
Bill Newsom
  Blue Cross and Blue Shield Association
Stuart L. Nightingale, M.D.
  Food and Drug Administration
Office for Medical Applications of Research
  National Institutes of Health
Seymour Perry, M.D.
  Institute for Health Policy Analysis
  Georgetown University Medical Center
Michael Pesko
  Food and Drug Administration
Richard Peto
  Oxford University
Thomas A. Preston, M.D.
  Seattle Public Health Hospital
Jeffrey L. Probstfield, M.D.
  Baylor-Methodist Lipid Research Clinic

Stanley Reiser, M.D.
  University of Texas Health Science Center
Michael A. Riddiough, D. Pharm.
  Pharmaceutical Manufacturers Association
Arnold Rosoff
  University of Pennsylvania
Leonard Saxe, Ph.D.
  Boston University
Daniel Siegel
  National Eye Institute
Richard Simon, Ph.D.
  National Cancer Institute
Michael L. Terrin, M.D.
  Maryland Medical Research Institute
William Vodra
  Arnold & Porter
Judith L. Wagner, Ph. D.
  Office of Technology Assessment
Byron Waksman
  Multiple Sclerosis Society
Norman W. Weissman, Ph.D.
  National Center for Health Services Research
Sankey Williams, M.D.
  University of Pennsylvania
Col. Phillip Winter, M.D.
  Department of Defense
Robert E. Wittes, M.D.
  National Cancer Institute
Donald Young, M.D.
  Health Care Financing Administration
Marvin Zelen, Ph.D.
  Harvard University School of Public Health

References

References

1. Amberson, J. B., McMahon, B.T., and Pinner, M., "A Clinical Trial of Sanocrysin in Pulmonary Tuberculosis," *Am. Rev. Tubercul.* 24:401-435, 1931.
2. American College of Surgeons Commission on Cancer, "Long-term and Short-term Surveys of Patterns of Care of the Female Breast in American College of Surgeons' Approved Cancer Programs," Report to Commission on Cancer Patient Care Research Committee, Oct. 22, 1982.
3. The Anturane Reinfarction Trial Policy Committee, "The Anturane Reinfarction Trial: Re-evaluation of Outcome," *N. Engl. J. Med.* 306(16): 1005-1008, Apr. 22, 1982.
4. The Anturane Reinfarction Trial Research Group, "Sulfinpyrazone in the Prevention of Sudden Death After Myocardial Infarction," *N. Engl. J. Med.* 302(5):250-256, Jan. 31, 1980.
5. Armitage, P., Bardelli, D., Galton, D. A. G., et al., *Methods and Impact of Controlled Therapeutic Trials in Cancer,* Part I. UICC Technical Report Series, vol. 36:132-133, Geneva, 1978.
6. Bailar, J. C., III, "Research Quality, Methodological Rigor, Citation Counts, and Impacts," *Am. J. Public Health* 72(10):1103-1104, 1982.
7. Banta, H. D., "RCTs and the Federal Government," *Contr. Clin. Tr.* 3(3):173-183, September 1982.
8. Banta, H. D., and Thacker, S., "Assessing the Costs and Benefits of Electronic Fetal Monitoring," *Obstet. & Gynecol. Sur.* 34(8):627-642, 1979.
9. Banta, H. D., Burns, A. K., and Behney, C. J., "Policy Implications of the Diffusion and Control of Medical Technology," *An. Am. Acad. Pol. & Soc. Sci.,* in press, 1983.
10. Barnes, B. A., "Discarded Operations: Surgical Innovation by Trial and Error," in *Costs, Risks, and Benefits of Surgery,* J. P. Bunker, B. A. Barnes, and F. Mosteller (eds.) (New York: Oxford University Press, 1977).
11. Barsamian, E. M., "The Rise and Fall of Internal Mammary Artery Ligation in the Treatment of Angina Pectoris and the Lessons Learned," in *Costs, Risks, and Benefits of Surgery,* J. P. Bunker, B. A. Barnes, and F. Mosteller (eds.) (New York: Oxford University Press, 1977).
12. Baum, M. L., Anish, D. S., Chalmers, T. C., et al., "A Survey of Clinical Trials of Antibiotic Prophylaxis in Colon Surgery: Evidence Against Further Use of No-Treatment Controls," *N. Engl. J. Med.* 305:795-799, 1981.
13. Beauchamp, T., and Childress, J. F. (eds.) *Principles of Biomedical Ethics* (New York: Oxford University Press, 1979). Cited in Lebacqz, 1980.
14. Begg, C., Elson, P., McFadden, E., et al., "Quality of Institutional Participation in Multicenter Trials," *N. Engl. J. Med.* 306(13):813, Apr. 1, 1982.
15. Bell, R., and Loop, J. W., "The Utility and Futility of Radiographic Skull Examination for Trauma," *N. Engl. J. Med.* 284:236, 1971.
16. Ben-Yishay, Y., Diller, L., Warga, C., et al., "The Alleviation of Cognitive and Functional Impairments in Senility by Hyperbaric Oxygenation Combined With Systematic Cuing," in *Fifth International Hyperbaric Conference,* vol. 1, W. G. Trapp, et al. (eds.) (Burnaby, Canada: Simon Fraser University Press, 1974).
17. Bloom, B., University of Pennsylvania, personal communication, Apr. 7, 1983.
18. Boissel, J. P., "Clinical Trials: Aims and Four Puzzling Points," in *Essais Controles Multicentres. Principes et Problemes,* J. P. Boissel and C. R. Klimt, (eds.) (Paris: Institute National de la Sante et de la Recherche en Medecine,1979), Cited in Marks and Rosenkrantz, 1982.
19. Boissel, J. P., et al., "Evaluation des Consequences des Resultats des Essais Cliniques Controles sur la Consommation Medicamenteuse," *Actions Thematiques Programmees, Institut National de la Sante et de la Recherche Medicale* 79-111, 1979.
20. Bonchek, L., "Are Randomized Trials Appropriate for Evaluating New Operations?" *N. Engl. J. Med.* 301(1):44-45, July 5, 1979.
21. Boruch, R., "On Common Contentions About Randomized Field Experiments," in *Experimental Testing of Public Policy: The Proceedings of the 1974 Social Science Research Council Conference on Social Experiments,* R. Boruch and H. Riecken (eds.) (Boulder, Colo.: Westview Press, 1975).
22. Boyle, E., Aparicio, A., Canosa, F., et al., "Hyperbaric Oxygen and Acetazolamide in the Treatment of Senile Cognitive Functions," in *Fifth International Hyperbaric Conference,* vol. 1, W. G. Trapp, et al. (eds.) (Burnaby, Canada: Simon Fraser University Press, 1974).
23. Breckenridge, A., "Should Every Survivor of a Heart Attack Be Given a Beta-blocker? Part II: Evidence From a Clinical Pharmacological Standpoint," *Br. Med. J.* 285:37-39, July 3, 1982.
24. Brown, B., "Issues in the Design of Clinical

Trials," Comments in Panel Discussion, J. C. Bailar, III, Chairman, Chapter 6 in Miké, 1983 (reference No. 161).

25. Brown, B., "Statistical Controversies in the Design of Clinical Trials—Some Personal Views," *Contr. Clin. Tr.* 1:13-27, 1980.
26. Buchwald, H., Fitch, L., and Moore, R., "Overview of Randomized Clinical Trials of Lipid Intervention for Atherosclerotic Cardiovascular Disease," *Contr. Clin.Tr.* 3(3):271-284, September 1982.
27. Bunker, J. P., and Fowles, J., "Model for an Institute for Health Care Evaluation," in *Strategies for Medical Technology Assessment,* OTA-H-181, Office of Technology Assessment, U.S. Congress, Washington, D.C., September 1982.
28. Bunker, J. P., Barnes, B. A., and Mosteller, F. (eds.) *Costs, Risks, and Benefits of Surgery* (New York: Oxford University Press, 1977).
29. Bunker, J. P., Hinkley, D., and McDermott, W., "Surgical Innovation and Its Evaluation," *Science* 200:937-941, May 26, 1978.
30. Bunker, J. P., Barnes, B. A., Mosteller, F., et al., "Summary, Conclusions, and Recommendations," in *Costs, Risks, and Benefits of Surgery,* J. P. Bunker, B. A. Barnes, and F. Mosteller (eds.) (New York: Oxford University Press, 1977).
31. Byar, D. B., "Why Data Bases Should Not Replace Randomized Clinical Trials," *Biometrics* 36:337-342, 1980.
32. Byar, D. B., Simon, R., Friedewald, W., et al., "Randomized Clinical Trials: Perspectives on Some Recent Ideas," *N. Engl. J. Med.* 295(2): 74-80, July 8, 1976.
33. Cancer Research Campaign Working Party in Breast Conservation, "Informed Consent: Ethical, Legal, and Medical Implications for Doctors and Patients Who Participate in Randomized Clinical Trials," *Br. Med. J.* 286:1117-1121, Apr. 2, 1983.
34. Canfield, C., Walter Reed Army Medical Center, personal communication, 1982.
35. Carbone, P. P., "Impact of the Cooperative Groups," in *Cancer Research: The Impact of the Cooperative Groups,* B. Hoogstraten (ed.) (New York: Masson Publishing, 1980).
36. Carpenter, W. T., Jr., Sadler, J. H., Light, P. D., et al., "The Therapeutic Efficacy of Hemodialysis in Schizophrenia," *N. Engl. J. Med.* 308(12), Mar. 24, 1983.
37. Cascinelli, N., Davis, H. L., Jr., Flamant, R., et al., *Methods and Impact of Controlled Therapeutic Trials in Cancer,* part II, UICC Technical Report Series (Geneva: International Union Against Cancer, 1981).
38. Cease, K., "Cooperative Clinical Trials and Continuing Medical Education," *N. Engl. J. Med.* 308(4):225-229, Feb. 10, 1983.
39. Chalmers, T. C., "The Impact of Controlled Trials on the Practice of Medicine," *Mt. Sinai J. Med.* 41:753-759, 1974.
40. Chalmers, T. C., "A Potpourri of RCT Topics," *Contr. Clin. Tr.* 3(3):285-298, September 1982.
41. Chalmers, T. C., "Randomization of the First Patient," *Med. Clin. North Am.* 59:1035-1038, 1975, cited in Chalmers, 1982.
42. Chalmers, T. C., "Scientific Use and Abuse of the Clinical Trial," paper presented at the Basel Liver Week, Basel, Switzerland, 1979.
43. Chalmers, T. C., "Who Will Fund Clinical Trials?" *The Sciences,* March 1982, pp. 6-8.
44. Chalmers, T. C., Berk, A. A., and Smith, H., "Early Mobilization and Discharge of Acute Myocardial Infarction Patients Apparently Saves Lives," *Clin. Res.* 27:508A, 1979.
45. Chalmers, T. C., Block, J., and Lee, S., "Controlled Studies in Clinical Cancer Research," *N. Engl. J. Med.* 287(2):75-78, July 13, 1972.
46. Chalmers, T. C., Smith, H., Jr., Blackburn, B., et al., Mt. Sinai School of Medicine, written personal communication, 1982.
47. Chalmers, T. C., Smith, H., Jr., Blackburn, B., et al., "A Method for Assessing the Quality of a Randomized Control Trial," *Contr. Clin. Tr.* 2:31-49. 1981.
48. Chalmers, T. C., Smith, H., Jr., Sacks, H., et al., Mt. Sinai School of Medicine, "The Technical Evaluation of the Clinical Literature," written communication, 1982.
49. Christensen, E., Juhl, E., and Tygstrup, N., "Treatment of Duodenal Ulcer:Randomized Clinical Trials of a Decade (1964-1974)," *Gastroenterology* 73:1170-1178, 1977.
50. Cochrane, A. L., *Effectiveness and Efficiency—Random Reflections on Health Services* (Abingdon: Nuffield Provincial Hospitals Trust, 1971). Cited in Marks and Rosenkrantz, 1982.
51. Cocco, A. E., and Cocco, D., "A Survey of Cimetidine Prescribing," *N. Engl. J. Med.* 304:1281, May 21, 1981.
52. Coleman, J. S., Katz, E., and Menzel, H., *Medical Innovation: A Diffusion Study* (Indianapolis: Bobbs-Merrill, 1966).
53. Colton, T., "Reporting on Methods in Clinical Trials," *N. Engl. J. Med.* 307(19):1219-1220, Nov. 4, 1982.
54. Comroe, J. H., Jr., and Dripps, R. D., "Scientific Basis for Support of Biomedical Science," *Science* 192:105, 1976.

55. Conn, H. O., and Blitzer, B. L., "Nonassociation of Adrenocorticosteroid Therapy and Peptic Ulcer," *N. Engl. J. Med.* 294:473-479, 1976.
56. Cook, T. D., and Campbell, D. T., *Quasi-Experimentation: Design and Analysis Issues for Field Settings* (Chicago: Rand McNally College Publishing Co., 1979).
57. Cooper, J. C., *The Philosophy of Evidence. 3rd Conference of the Philosophy and Technology of Drug Assessment* (Washington, D.C.: Smithsonian Institution, 1975), cited in Marks and Rosenkrantz, 1982.
58. Coordinating Center Models Project, "A Study of Coordinating Centers in Multicenter Clinical Trials," prepared by the Coordinating Center Models Project Research Group, University of Maryland School of Medicine, Baltimore, Md., Apr. 1, 1980, cited in Probsfield, et al., 1983.
59. Davis, H. L., Durant, J. R., and Holland J. F., "Interrelationships: the Groups, the NCI and Other Governmental Agencies," in *Cancer Research: The Impact of the Cooperative Groups,* B. Hoogstraten (ed.) (New York: Masson Publishing, 1980).
60. De Dombal, F., "Quality of Institutional Participation in Multicenter Trials," *N. Engl. J. Med.* 306(13):813, Apr. 1, 1982.
61. DerSimonian, R., Charette, L., McPeek, B., et al., "Reporting on Methods in Clinical Trials," *N. Engl. J. Med.* 306(22):1332-1337, June 3, 1982.
62. DerSimonian, R., Charette, L., McPeek, B., et al., "Reporting on Methods in Clinical Trials," *N. Engl. J. Med.* 307(19):1220, Nov. 4, 1982.
63. Dowling, H., *Medicines for Man* (New York: Alfred A. Knopf, 1970).
64. Drummond, M., and Mooney, G., "Essentials of Health Economics, Part V—Assessing the Costs and Benefits of Treatment Alternatives," *Br. Med. J.* 285:1561-1563, Nov. 27, 1982.
65. The Eastern Cooperative Oncology Group Writing Committee, "Participation of Community Hospitals in Clinical Trials," *N. Engl. J. Med.* 306(18):1076-1080, May 6, 1982.
66. Edwards, A., and Hart, G., "Hyperbaric Oxygenation and the Cognitive Functioning of the Aged," *J. Am. Geriatr. Soc.* 22:376, 1974.
67. Eisenberg, J., and Williams, S., "Cost Containment and Changing Physicians' Practice Behavior," *J.A.M.A.* 246(19):2195-2201, Nov. 13, 1981.
68. Elkeles, A., "Informed Consent in Clinical Trials," *Lancet* 1:1189, 1982.
69. European Coronary Surgery Study Group, "Long-Term Results of Prospective Randomized Study of Coronary Artery Bypass Surgery in Stable Angina Pectoris," *Lancet* 1: 1174-1180, Nov. 27, 1982.
70. Fineberg, H. V., "Gastric Freezing—A Study of Diffusion of a Medical Innovation," App. D in *Medical Technology and the Health Care System* (Washington, D.C.: National Academy of Sciences, 1979), cited in Wortman, 1981.
71. Fineberg, H. V., and Hiatt, H. H., "Evaluation of Medical Practices: The Case for Technology Assessment," *N. Engl. J. Med.* 301:1086-1091, 1979, cited in Marks and Rosenkrantz, 1982.
72. Finkelstein, S. N., and Gilbert, D. L., "Scientific Evidence and the Abandonment of Medical Technology: A Study of Eight Drugs," unpublished, March 1983.
73. Fisher, B., "The Interdependence of Laboratory and Clinical Research in the Study of Metastases," presented at the 1983 Annual Symposium on Fundamental Cancer Research, Houston, Tex., Feb. 28-Mar. 3, 1983.
74. Fisher, L., and Kennedy, J., "Randomized Surgical Clinical Trials for Treatment of Coronary Artery Disease," *Contr. Clin. Tr.* 3(3):235-258, September 1982.
75. Fletcher, R., and Fletcher, S., "Clinical Research in General Medical Journals—A 30-Year Perspective," *N. Engl. J. Med.* 301(4):180-183, July 1979.
76. Fox, R., "The Human Condition of Health Professionals," lecture presented at the University of New Hampshire, 1979.
77. Fox, R., "Situations Creating Ethical Stress," *Med. J. Australia* 1:162-163, 1981.
78. Fredrickson, D., "Biomedical Research in the 1980's," *N. Engl. J. Med.* 304(9):509-517, Feb. 26, 1981.
79. Fredrickson, D., "The Field Trial: Some Thoughts on the Indispensable Ordeal," *Bull. N. Y. Acad. Med.* 44:985-993, 1968, cited in Marks and Rosenkrantz, 1982.
80. Freidson, E., *Profession of Medicine* (New York: Dodd, Mead & Co., 1970).
81. Freis, E. D., "Should Mild Hypertension Be Treated?" *N. Engl. J. Med.* 307(5):306-309, Feb. 26, 1981.
82. Friedewald, W., and Schoenberger, J., "Overview of Recent Clinical and Methodological Advances From Clinical Trials of Cardiovascular Disease," *Contr. Clin. Tr.* 3(3):259-270, September 1982.
83. Friedman, L., and Furberg, C., "Community Hospitals in Clinical Trials"(correspondence), *N. Engl. J. Med.* 307(15):958, Oct. 7, 1982.
84. Friedman, L., Furberg, C., and deMets, D., *Fundamentals of Clinical Trials* (Boston: Wright-PSG, 1981).

85. Fuerst, M. L., "Is Screening for Early Lung Cancer Worthwhile?" *J.A.M.A.* 249(12):1537-1539, Mar. 25, 1983.
86. Garnier, H., Flamant, R., and Fohanno, C., "Assessment of the Role of Randomized Clinical Trials in Establishing Treatment Policies," *Contr. Clin. Tr.* 3(3):227-234, September 1982.
87. Gehan, E., "Progress of Therapy in Acute Leukemia 1948-1981: Randomized Versus Nonrandomized Clinical Trials," *Contr. Clin. Tr.* 3(3): 199-208, September 1982.
88. Gehan, E., and Freireich, E., "Non-randomized Controls in Cancer Clinical Trials," *N. Engl. J. of Med.* 290(4):198-203, Jan. 24, 1974.
89. Gilbert, J., McPeek, B., and Mosteller, F., "Progress in Surgery and Anesthesia: Benefits and Risks of Innovative Therapy," in *Costs, Risks, and Benefits of Surgery,* J. P. Bunker, B. A. Barnes, and F. Mosteller (eds.) (New York: Oxford University Press, 1977).
90. Gilchrist, V., "Dissemination of Information on Hypertension," *J.A.M.A.* 247(7):977, Feb. 19, 1982.
91. Glaser, E. M., "Volunteers, Controls, Placebos, and Questionnaires in Clinical Trials," in *Medical Surveys and Clinical Trials. Some Methods and Application of Group Research in Medicine,* L. J. Witts (ed.) (London: Oxford University Press, 1959), cited in Marks and Rosenkrantz, 1982.
92. Glass, G. V., "Primary, Secondary, and Meta-Analysis of Research," *Educ. Res.* 5:3, 1976.
93. Glass, G. V., and Smith, M. L., "Meta-Analysis of Research on Class Size and Achievement," *Educ. Eval. Policy Analysis* 1:2, 1979.
94. Glore, J., "Seventeenth Century Medicine," in Program Book for Moliere's "The Imaginary Invalid," Arena Stage, Washington, D.C., Jan. 21, 1983.
95. Goldfarb, A., Hochstadt, N., Jacobson, J., et al., "Hyperbaric Oxygen Treatment of Organic Mental Syndrome in Aged Persons," *J. Gerontol.* 27:212, 1972.
96. Gordon, G., and Fisher, G. L. (eds.) *The Diffusion of Medical Technology* (Cambridge, Mass.: Ballinger Publishing Co., 1975).
97. Gordon, R. S., National Institutes of Health, personal communication, March 1983.
98. Gould, B., Mann, S., Davies, A., et al., "Does Placebo Lower Blood-pressure?" *Lancet* 2: 1377-1381, Dec. 19, 1981.
99. Green, S. B., "Patient Heterogeneity and the Need for Randomized Clinical Trials," *Contr. Clin. Tr.* 3(3):189-198, September 1982.
100. Greer, A. L., "Advances in the Study of Diffusion and Innovation in Health Care Organizations," *Mil. Mem. Fund Qrly./Health Soc.* 55:505, 1977.
101. Greer, A. L., "Deus Ex Machina: Physicians in the Adoption of Hospital Medical Technology," Urban Research Center, Milwaukee, Wis., unpublished manuscript,1981.
102. Griner, P. F., and Glaser, R. J., "Misuse of Laboratory Tests and Diagnostic Procedures," *N. Engl. J. Med.* 307(21):1336-1338, Nov. 18, 1982.
103. Haines, S. J., "Randomized Clinical Trials in Neurosurgery," *J. Neurosurg.* 12:259-264, 1983.
104. Haines, S. J., "Randomized Clinical Trials in the Evaluation of Surgical Innovation," *J. Neurosurg.* 51:5-11, 1979.
105. Haines, S. J., "Should Antifibrinolytic Therapy After Subarachnoid Hemorrhage Be Abandoned?" typescript, 1983.
106. Hampton, J., "Should Every Survivor of a Heart Attack Be Given a Beta-Blocker? Part I: Evidence From Clinical Trials," *Br. Med. J.* 285:33-36, July 3, 1982.
107. Hemminki, E., "Quality of Clinical Trials—A Concern of Three Decades," *Meth. Inform. Med.* 21:81-85, 1982.
108. Hemminki, E., "Quality of Reports of Clinical Trials Submitted by the Drug Industry to the Finnish and Swedish Control Authorities," *Eur. J. Chm. Pharmacol.* 19:157-165, 1981.
109. Hiatt, H. H., "Lessons of the Coronary-Bypass Debate," *N. Engl. J. Med.* 297(26):1462-1464, Dec. 29, 1977.
110. Hiatt, H. H., "Protecting the Medical Commons: Who Is Responsible?" *N. Engl. J. Med.* 293(5): 235-241, July 31, 1975.
111. Hill, A. B., "Aims and Ethics," *Conference on Controlled Clinical Trials—Vienna 1959* (Oxford: Blackwell Scientific Publishers, 1960), cited in Marks and Rosenkrantz, 1982.
112. Hill, A. B., "Assessment of Therapeutic Trials," *Transactions of the Medical Society of London* 68:128-136, 1953, cited in Marks, 1981.
113. Hill, A. B., *Principles of Medical Statistics,* 3d ed. (London: The Lancet, Ltd., 1945), cited in Marks and Rosenkrantz, 1982.
114. Hoekelman, R. A., "The Impact of Ambulatory Pediatric Research on Changing Behavior," *Am. J. Dis. Child.* 132:1079-1081, 1978.
115. Imai, Y., "Hyperbaric Oxygen (OHP) Therapy in Memory Disturbances," in *Fifth International Hyperbaric Conference,* vol. 1, W. G. Trapp, et al. (eds.) (Burnaby, Canada: Simon Fraser University Press, 1974).
116. Jacobs, E., Alvis, H., and Small, S., "Hyperoxygenation: A Central Nervous System Activator?" *J. Geriatr. Psyc.* 5:107, 1972.

117. Jacobs, E., Winter, P., Alvis, H., et al., "Hyperoxygenation Effects on Cognitive Functioning in the Aged," *N. Engl. J. Med.* 281:753, 1969.
118. Johnson, W., "Did Too Many Nonclinicians Spoil $115-million MRFIT?" *Med. World N.* 23(25):140, Dec. 6, 1982.
119. Kaae, S., and Johansen, H., "Simple Mastectomy Plus Postoperative Irradiation by the Method of McWhirter for Mammary Carcinoma," *Prog. Clin. Cancer* 1:453, 1965, cited in McPherson and Fox, 1977.
120. Kaluzny, A., Barhyte, D. D., Reader, G. C., et al., "Health Systems," in *The Diffusion of Medical Technology,* G. Gordon and L. Fisher (eds.) (Cambridge, Mass.: Ballinger Publishing Co., 1975).
121. Kaplan, N., "Therapy for Mild Hypertension," *J.A.M.A.* 249(3):365-367, Jan. 21, 1983.
122. Kessner, D., "Diffusion of New Medical Information," *Am. J. Public Health* 71(4):367-368, April 1981.
123. Klimt, C. R., "The Conduct and Principles of Randomized Clinical Trials," *Contr. Clin. Tr.* 1(4):283-293, 1981.
124. Klimt, C. R., "Terminating a Long-Term Clinical Trial," *Contr. Clin. Tr.* 1(4):319-325, 1981.
125. Kopanda, R., Alcohol, Drug Abuse and Mental Health Administration, personal communication, 1983.
126. Korenbrot, C., Flood, A. B., Higgins, M., et al., *Elective Hysterectomy: Costs, Risks, and Benefits,* OTA-BP-H-9(15), Case Study #5 of OTA Health Technology Case Study Series, prepared for the Office of Technology Assessment, U.S. Congress, Washington, D.C., October 1981.

126a. *Lancet,* "Aspirin After Myocardial Infarction" (editorial), *Lancet* 1:1172-1173, May 31, 1980.

127. *Lancet,* "Coronary Disease and Multiple Risk Factor Intervention" (editorial), *Lancet* 1:1395, June 19, 1982.
128. *Lancet,* "Long-Term and Short-Term Beta-blockade After Myocardial Infarction" (editorial), *Lancet* 1:1159-1162, May 22, 1982.
129. *Lancet,* "Recruiting for Multicentre Trials" (in Notes and News), *Lancet* 1:78, 1983.
130. *Lancet,* "Trials of Coronary Heart Disease Prevention" (editorial), *Lancet* 1:803-804, Oct. 9, 1982.
131. Lasagna, L., "Historical Controls: The Practitioner's Clinical Trials," *N. Engl. J. Med.* 307(21): 1339-1340, Nov. 18, 1982.
132. Lebacqz, K., "Controlled Clinical Trials: Some Ethical Issues," *Contr. Clin. Tr.* 1:29-36, 1980.
133. Lebacqz, K., and Levine, R. J., "Respect for Persons and Informed Consent to Participate in Research," *Clin. Res.* 25(3):101-107, no date. Cited in Lebacqz, K., 1980.
134. Levy, R., and Sondik, E., "Decision-making in Planning Large-scale Comparative Studies," *N. Y. Acad. Sci.* 304:441-457, 1978.
135. Light, R. J., and Smith, P. V., "Accumulating Evidence: Procedures for Resolving Contradictions Among Research Studies," *Harvard Educ. Rev.* 41:429,1976.
136. Lloyd, C., "The Introduction of Lemon Juice as a Cure for Scurvy," *Bull. Hist. Med.* 35:123-132, 1961.
137. Love, J., "Drugs and Operations—Some Important Differences," *J.A.M.A.* 232(1):37-38, Apr. 7, 1975.
138. Lundberg, G. D., "MRFIT and the Goals of the Journal" (editorial), *J.A.M.A.* 248(12):1501, 1982.
139. Management Committee of the Australian Therapeutic Trial in Mild Hypertension, "Untreated Mild Hypertension," *Lancet* 1:185-191, Jan. 23, 1982.
140. Marieskind, H., *An Evaluation of Cesarean Section in the United States* (Washington, D.C.: Office of Assistant Secretary for Planning and Evaluation/Health, DHEW, 1979).
141. Marks, H., Department of Social Medicine and Health Policy, Harvard Medical School, personal communication, Apr. 1, 1983.
142. Marks, H., "Medical Science and Clinical Ambiguity: The Role of Randomized Clinical Trials in Therapeutic Controversies," draft typescript, Harvard School of Public Health, November 1981.
143. Marks, H., and Rosenkrantz, B., "Clinical Trials and Clinical Judgement: The Role of Evidence in Therapeutic Choice," typescript, Harvard Medical School and Harvard School of Public Health, August 1982.
144. Marwick, C., " 'Mild' Hypertension: The Gray Zone Gets More Confusing," *Med. World N.* 23(26):66-85, Dec. 20, 1982.
145. Mayers, A. E., National Center for Health Services Research, personal communication, Apr. 13, 1983.
146. McAlister, N., "Should We Treat 'Mild' Hypertension?" *J.A.M.A.* 249(3):379-383, Jan. 21, 1983.
147. McDowell, F. H., "The Extracranial/Intracranial Bypass Study" (editorial), *Stroke* 8(5):545, 1977.
148. McGhan, W. F., Smith, W. E., and Adams, D. W., "A Randomized Trial Comparing Pharmacists and Technicians as Dispensers of Prescriptions for Ambulatory Patients," *Med. Care* 21(4):445-453, April 1983.

149. McGrady, G. A., "The Controlled Clinical Trial and Decision Making in Family Practice," *J. Fam. Prac.* 14(4):739-744, 1982.
150. McKinlay, J. B., "From 'Promising Report' to 'Standard Procedure': Seven Stages in the Career of a Medical Innovation," *Mil. Mem. Fund Qtrly./Health & Soc.* 59(3):374, summer 1981.
151. McKinlay, S. M., "Experimentation in Human Populations," *Mil. Mem. Fund Qtrly./Health & Soc.* 59(3):308-323, summer 1981.
152. McPeek, B., Gilbert, J., and Mosteller, F., "The End Result: Quality of Life," in *Costs, Risks, and Benefits of Surgery,* J. P. Bunker, B. A. Barnes and F. Mosteller (eds.) (New York: Oxford University Press, 1977).
153. McPherson, K., and Fox, M., "Treatment of Breast Cancer," in *Costs, Risks, and Benefits of Surgery,* J. P. Bunker, B. A. Barnes, and F. Mosteller (eds.) (New York: Oxford University Press, 1977).
154. McQuillen, M., University of Minnesota, personal communication, 1982.
155. Mechanic, D., "The Growth of Medical Technology and Bureaucracy: Implication for Medical Care," *Mil. Mem. Fund Qtrly./Health Soc.* 55:61, 1977.
156. *Medical World News,* "Doctors Don't See Power of Drug Ads," *Med. World N.* 24:19-20, Sept. 13, 1982.
157. Meier, P., "Stratification in the Design of a Clinical Trial," *Contr. Clin.Tr.* 1:355-361, May 1981.
158. Meinert, C., "Funding for Clinical Trials," *Contr. Clin. Tr.* 3(3):165-171, September 1982.
159. Meinert, C., "Toward More Definitive Clinical Trials," *Contr. Clin. Tr.* 1:249-261, 1980.
160. Miao, L., "Gastric Freezing: An Example of the Evaluation of Medical Therapy by Randomized Clinical Trials," in *Costs, Risks, and Benefits of Surgery,* J. P. Bunker, B. A. Barnes, and F. Mosteller (eds.) (New York: Oxford University Press, 1977).
161. Miké, V., and Stanley, K. E. (eds.) *Statistics in Medical Research* (New York: John Wiley & Sons, 1983).
162. Miller, A. B., and Bulbrook, R. D., "Screening, Detection, and Diagnosis of Breast Cancer," *Lancet* 1:1109-1111, 1982.
163. Moskowitz, G., Chalmers, T. C., Sacks, H., et al., "Deficiencies of Controlled Trials of Alcohol Withdrawal," *Alcoholism Clin. Exp. Res.* 5:162, 1981.
164. Mosteller, F., and Weinstein, M., "Toward Evaluating the Cost-Effectiveness of Medical and Social Experiments," typescript, Harvard School of Public Health, 1982.
165. Mosteller, F., Gilbert, J., and McPeek, B., "Reporting Standards and Research Strategies for Controlled Trials," *Contr. Clin. Tr.* 1:37-58, 1980.
166. Multiple Risk Factor Intervention Trial Research Group, "Multiple Risk Factor Intervention Trial: Risk Factor Changes and Mortality Results," *J.A.M.A.* 248(12):1465-1501, Sept. 24, 1982.
167. National Academy of Sciences, Committee on Technology and Health Care, *Medical Technology and the Health Care System* (Washington, D.C.: NAS, 1979).
168. Neuhauser, D., Cohen, D., and Jones, P., "The Metro Firm Trials: An Innovative Approach to Ongoing Randomized Clinical Trials," typescript, May 17, 1982.
169. Newsom, B., The Blue Cross/Blue Shield Association, personal communication, April 1983.
170. Nickson, J., and Glicksman, A., "Clinical Trials in Radiation Therapy: An Application of Statistical Methods to Clinical Research," *Am. J. Roentgenol.* 96(1):236-240, January 1966.
171. Nightingale, S. L., "Drug Regulation and Policy Formulation," *Mil. Mem. Fund Qtrly./Health & Soc.* 59(3):412-444, summer 1981.
172. Orloff, M. J., et al., "Contributions of Surgical Research to Health Care," 1945-1970 Report of the Research Subcommittee of the Study on Surgical Services for the United States (Washington, D.C.: American College of Surgeons and American Surgical Association, 1976). Cited in Bunker, Barnes, and Mosteller, 1977.
173. Palva, I., "Quality of Institutional Participation in Multicenter Trials," *N. Engl. J. Med.* 306(13): 814, Apr. 1, 1982.
174. Paradise, J., and Bluestone, C., "Toward Rational Indications for Tonsil and Adenoid Surgery," *Hospital Practice* 11:79-87, February 1976.
175. Patulin Clinical Trials Committee, "Clinical Trial of Patulin in the Common Cold," *Lancet* 2: 373-375, 1944.
176. Perrow, C., "Hospitals: Technology, Structure, and Goals," in *Handbook for Organizations,* J. March (ed.) (Chicago: Rand McNally, 1965).
177. Perry, S., "The Brief Life of the National Center for Health Care Technology," *N. Engl. J. Med.* 307(17):1095-1100, Oct. 21, 1982.
178. Perry, S., "Technology Assessment Proposed," *Health Affairs* 1(4):123-128, fall 1982.
180. Peto, R., Pike, M., Armitage, P., et al., "Design and Analysis of Randomized Clinical Trials Requiring Prolonged Observation of Each Patient, Part I," *Br. J. Cancer* 34:585-607, August 1976.
181. Peto, R., Pike, M., Armitage, P., et al., "Design and Analysis of Randomized Clinical Trials Re-

quiring Prolonged Observation of Each Patient, Part II," *Br. J. Cancer* 35:1-29, August 1976.

182. Pharmaceutical Manufacturers Association, *PMA 1979-1980 Annual Survey Report* (Washington, D.C.: PMA, no date).
183. Pickering, T., "Treatment of Mild Hypertension and the Reduction of Cardiovascular Mortality: The 'Of or By' Dilemma," *J.A.M.A.* 249(3): 399-400, Jan. 21, 1983.
184. Pocock, S. J., and Lagakos, S. W., "Practical Experience of Randomization in Cancer Trials: An International Survey," in *Evaluation of Methods of Treatment and Diagnostic Procedures in Cancer,* T. B. Grage, G. A. Higgins, S. W. Lagakos, et al. (eds.) (Geneva, Switzerland: International Union Against Cancer, 1982).
185. Probstfield, J. L., Russell, M. L., Silvers, A., et al., "The Relevance of Applying Clinical Trials Methods to the Practice of Clinical Medicine," typescript, 1983.
186. Raskin, A., Gershon, S., Crook, T., et al., "The Effects of Hyperbaric and Normobaric Oxygen on Cognitive Impairment in the Elderly," *Arch. G. Psych.* 35:50, 1978.
187. Redisch, M., "Hospital Inflationary Mechanisms," presented at the meeting of the Western Economic Association, Las Vegas, Nev., 1974.
188. Redmond, C., and Fisher, B., "The Prospective Randomized Controlled Clinical Trial: Its Design and Other Considerations," prepared under contract by the National Surgical Adjuvant Breast and Bowel Project, University of Pittsburgh for the American Cancer Society, contract #NO1-CB-23876, to be published in *Surgical Oncology,* McGraw Hill Book Co., in press.
189. Riegelman, R., George Washington University Medical Center, personal communication, April 1983.
190. Rockette, H., Redmond, C., Fisher, B., et al., "Impact of Randomized Clinical Trials on Therapy of Primary Breast Cancer: The NSABP Overview," *Contr. Clin. Tr.* 3(3):209-226, September 1982.
191. Roemer, M., and Shain, M., *Hospital Utilization Under Insurance* (Chicago: American Hospital Association, 1959).
192. Rogers, E. M., and Shoemaker, F. F., *Communication of Innovations: A Cross-Cultural Approach* (New York: The Free Press, 1971).
193. Rose, G., "Should Every Survivor of a Heart Attack Be Given a Beta-blocker? Part III: Some Conclusions," *Br. Med. J.* 285:39-40, July 3, 1982.
194. Ruby, G., Office of Technology Assessment, U.S. Congress, "Coverage, Medical Technology Evaluation and Medicare Costs," draft typescript, 1983.
195. Russell, L., *Technology in Hospitals: Medical Advances and Their Diffusion* (Washington, D.C.: The Brookings Institution, 1979).
196. Sackett, D. L., and Gent, M., "Controversy in Counting and Attributing Events in Clinical Trials," *N. Engl. J. Med.* 301:1410-1412, 1979.
197. Sacks, H., Chalmers, T., and Smith, H., "Randomized Versus Historical Controls for Clinical Trials," *A. J. M.* 72:233-240, February 1982.
198. Saxe, L., and Fine, M., *Social Experiments: Methods for Design and Evaluation* (Beverly Hills, Calif.: Sage Publications, Inc., 1981).
199. Schade, R. R., and Donaldson, R. M., Jr., "How Physicians Use Cimetidine," *N. Engl. J. Med.* 304:1281-1284, May 21, 1981.
200. Schifrin, L. G., and Tayan, J. R., "The Drug Lag: An Interpretive Review of the Literature," *Int. J. Health Serv.* 7:359, 1977.
201. Schneyer, S., Landefeld, J., and Sandifer, F., "Biomedical Research and Illness: 1900-1979," *Milb. Mem. Fund Qrtrly./Health & Soc.* 59(1): 44-58, 1981.
202. Schroeder, S. A., and Showstack, J. A., "Financial Incentives To Perform Medical Procedures and Laboratory Tests: Illustrative Models of Office Practice," *Med. Care* 16:189, 1978.
203. Schroeder, S. A., and Showstack, J. A., "The Dynamics of Medical Technology Use: Analysis and Policy Options," in *Medical Technology: The Culprit Behind Health Care Costs,* S. Altman and R. Blendon (eds.) (Hyattsville, Md.: National Center for Health Services and Bureau of Health Planning, 1979).
204. Shapiro, S., Strax, P., and Venet, L., "Periodic Breast Cancer Screening in Reducing Mortality From Breast Cancer," *J.A.M.A.* 215(11):1777-1785, Mar. 15,1971.
205. Showstack, J. A., Schroeder, S. A., and Steinberg, H. R., "Evaluating the Costs and Benefits of a Diagnostic Technology: The Case of Upper Gastrointestinal Endoscopy," *Med. Care* 19(5): 498, May 1981.
206. Silverman, M., and Lee, P. R., *Pills, Profits, and Politics* (Berkeley, Calif.: University of California Press, 1974).
207. Silverman, B., Shareck, E. P., and Ambroz, A., "Duodenal Ulcer Treatment: An Analysis of the Randomized Control Trials," *Gastroenterology* 74:1174, 1978.
208. Simon, R., National Cancer Institute, personal communication, April 1983.
209. Society for Clinical Trials, Inc., "Society for Clin-

ical Trials—By-Laws," *Contr. Clin. Tr.* 1(1):83-89, 1980.

210. Spodick, D., "Randomized Controlled Clinical Trials," *J.A.M.A.* 247(16):2258-2260, Apr. 23, 1982.
211. Staquet, M. (ed.), *Randomized Trials in Cancer: A Critical Review by Sites,* Monograph Series of the European Organization for Research on Treatment of Cancer, vol. 4 (New York: Raven Press, 1978).
212. Stevens, R., *American Medicine and the Public Interest* (New Haven, Conn.: Yale University Press, 1971).
213. Stross, J., and Harlan, W., "The Dissemination of New Medical Information," *J.A.M.A.* 241(24): 2622-2624, June 15, 1979.
214. Stross, J., and Harlan, W., "Dissemination of Relevant Information on Hypertension," *J.A.M.A.* 246(4):360-362, July 1981.
215. Sylvester, R. J., Pinedo, H. M., DePauw, M., et al., "Quality of Institutional Participation in Multicenter Trials," *N. Engl. J. Med.* 305:852-855, 1981.
216. Tabar, L., and Gad, A., "Screening for Breast Cancer: The Swedish Trial," *Radiology* 138: 219-222, January 1981.
217. Tanon, C. P., and Rogers, E., *Diffusion of Medical Technology,* G. Gordon and L. Fisher (eds.) (Cambridge, Mass.: Ballinger Publishing Co., 1975).
218. Taves, D. R., "Minimization: A New Method of Assigning Patients to Treatment and Control Groups," *Clin. Pharm. & Ther.* 15(5):443-453, May 1974.
219. Temin, P., *Taking Your Medicine: Drug Regulation in the United States* (Cambridge, Mass.: Harvard University Press, 1980).
220. Temple, R., and Pledger, G. W., "The FDA's Critique of the Anturane Reinfarction Trial," *N. Engl. J. Med.* 303(5):1488-1492, 1980.
221. Thomas, P., Novak, J., Knowlton, A., et al., "Quality of Institutional Participation in Multicenter Trials," *N. Engl. J. Med.* 306(13):814, Apr. 1, 1982.
222. Thompson, L., Davis, G., Obrist, W., et al., "Effects of Hyperbaric Oxygen on Behavioral and Physiological Measures in Elderly Demented Patients," *J. Gerontol.* 31:23, 1976.
223. Towery, O., and Perry, S., "The Scientific Basis for Coverage Decisions by Third-Party Payers," *J.A.M.A.* 245(1):59-61, Jan. 2, 1982.
224. Tygstrup, N., Lachin, J. M., and Juhl, E., *The Randomized Clinical Trial and Therapeutic Decisions* (New York: Marcel Dekker, 1982).
225. U.S. Congress, Office of Technology Assessment, *Assessing the Efficacy and Safety of Medical Technologies,* OTA-H-75, Washington, D.C., September 1978.
226. U.S. Congress, Office of Technology Assessment, *Assessment of Technologies for Determining Cancer Risks From the Environment,* OTA-H-138, Washington, D.C., June 1981.
227. U.S. Congress, Office of Technology Assessment, *Development of Medical Technology: Opportunities for Assessment* OTA-H-34, Washington, D.C., August 1976.
228. U.S. Congress, Office of Technology Assessment, *Forecasts of Physician Supply and Requirements,* OTA-H-113, Washington, D.C., April 1980.
229. U.S. Congress, Office of Technology Assessment, *The Implications of Cost-Effectiveness Analysis of Medical Technology,* OTA-H-126, Washington, D.C., August 1980.
230. U.S. Congress, Office of Technology Assessment, *Medical Technology Under Proposals To Increase Competition in Health Care,* OTA-H-190, Washington, D.C., October 1982.
231. U.S. Congress, Office of Technology Assessment, *Patent-Term Extension and the Pharmaceutical Industry,* OTA-CIT-143, Washington, D.C., August 1981.
232. U.S. Congress, Office of Technology Assessment, *Postmarketing Surveillance of Prescription Drugs,* OTA-H-189, Washington, D.C., November 1982.
233. U.S. Congress, Office of Technology Assessment, *A Review of Selected Federal Vaccine and Immunization Policies Based on Case Studies of Pneumococcal Vaccine,* OTA-H-96, Washington, D.C., September 1979.
234. U.S. Congress, Office of Technology Assessment, *Strategies for Medical Technology Assessment,* OTA-H-181, Washington, D.C., September 1982.
235. U.S. Congress, Office of Technology Assessment, *Technology Transfer at the National Institutes of Health: A Technical Memorandum,* OTA-TM-H-10, Washington, D.C., March 1982.
236. U.S. Department of Health and Human Services, Food and Drug Administration, "FDA Interim Trade Name Index of All Prescription Drugs in the Drug Efficacy Study—Cumulative Up to Nov. 1, 1982. Unfinished Products Only," computer printout.
237. U.S. Department of Health and Human Services, Office of Health Research, Statistics, and Technology, "PHS Assessment Policy and Processes," draft, Rockville, Md., Sept. 23, 1982.

238. U.S. Department of Health and Human Services, Public Health Service, *Compilation of Experimental Cancer Therapy Protocol Summaries,* NIH publication No. 82-1116 (Rockville, Md.: DHHS, April 1982).

239. U.S. Department of Health and Human Services, National Institutes of Health, *National Institutes of Health Research Plan, FY 1983-85,* 1981.

240. U.S. Department of Health and Human Services, National Institutes of Health, National Heart, Lung, and Blood Institute, *Clinical Trials Reference Document,* typescript, February 1982.

241. U.S. Department of Health, Education and Welfare, *Report of the Secretary's Commission on Medical Malpractice* (Washington, D.C.: DHEW, 1973).

242. U.S. Department of Health, Education, and Welfare, National Institutes of Health, *NIH Inventory of Clinical Trials: Fiscal Year 1975,* (Bethesda, Md.: NIH, 1975).

243. Veterans Administration, Department of Medicine and Surgery, *Focus: VA Research and Development—Hypertension* (pamphlet) (Washington, D.C.: Department of Surgery, May 1981).

244. Wagemaker, H., Jr., and Cade, R., "The Use of Hemodialysis in Chronic Schizophrenia," *Am. J. Psych.* 134:684, 1977.

245. Wangensteen, O. H., Peter, E. T., Nicoloff, D. M., et al., "Achieving 'Physiological Gastrectomy' by Gastric Freezing: A Preliminary Report of an Experimental and Clinical Study," *J.A.M.A.* 180:439, 1962. Cited in Miao, 1977.

246. Warner, K., "A 'Desperation-Reaction' Model of Medical Diffusion," *Health Serv. Res.* 10(4): 369-383, winter 1975.

247. Warner, K., "Effects of Hospital Cost Containment on the Development and Use of Medical Technology," *Mil. Mem. Fund Qtrly./Health & Soc.* 56:187, 1978.

248. Weingrod, K. S., and Neuhauser, D., *Surgery for Breast Cancer,* OTA-BP-H-9(17), Case Study #17 of OTA Health Technology Case Study Series, Office of Technology Assessment, U.S. Congress, Washington, D.C., October 1981.

249. Wennberg, J., and Gittelsohn, A., "Health Care Delivery in Maine, Patterns of Use of Common Surgical Procedures," *J. Maine Med. Assoc.* 66:123, 1975.

250. World Health Organization, *International Classification of Diseases, 1975 Revision* (Geneva: World Health Organization, 1977).

251. Wortman, P. M. (ed.), "Randomized Clinical Trials," in *Methods for Evaluating Health Services* (Beverly Hills, Calif.: Sage Publications, 1981).

252. Wortman, P. M., and Saxe, L., "Assessment of Medical Technology: Methodological Considerations," in *Strategies for Medical Technology Assessment,* OTA-H-181, Office of Technology Assessment, U.S. Congress, Washington, D.C., September 1982.

253. Wortman, P. M., and Yeaton, W., "Synthesis of Results in Controlled Trials of Coronary Artery Bypass Graft Surgery," typescript, Center for Research on the Utilization of Scientific Knowledge, University of Michigan, 1982.

254. Young, D., Health Care Financing Administration, personal communication, November 1982.

255. Zelen, M., Sidney Farber Cancer Institute, Boston, personal communication, January 1982.

256. Zelen, M., "Play the Winner Rule and the Controlled Clinical Trial," *J. Am. Stat. Assoc.* 64:131-146, 1969.

257. Zelen, M., "Strategy and Alternate Randomized Designs in Cancer Clinical Trials," *Cancer Treatment Rpts.* 66(5):1095-1100, 1982.

258. Zelen, M., Gehan, E., and Glidewell, O., "Biostatistics," in *Cancer Research: The Impact of the Cooperative Groups,* B. Hoogstraten (ed.) (New York: Masson Publishing, 1980).

259. Zubrod, C., "Clinical Trials in Cancer Patients: An Introduction," *Contr. Clin. Tr.* 3(3):185-188, September 1982.

Office of Technology Assessment

The Office of Technology Assessment (OTA) was created in 1972 as an analytical arm of Congress. OTA's basic function is to help legislative policymakers anticipate and plan for the consequences of technological changes and to examine the many ways, expected and unexpected, in which technology affects people's lives. The assessment of technology calls for exploration of the physical, biological, economic, social, and political impacts that can result from applications of scientific knowledge. OTA provides Congress with independent and timely information about the potential effects—both beneficial and harmful—of technological applications.

Requests for studies are made by chairmen of standing committees of the House of Representatives or Senate; by the Technology Assessment Board, the governing body of OTA; or by the Director of OTA in consultation with the Board.

The Technology Assessment Board is composed of six members of the House, six members of the Senate, and the OTA Director, who is a nonvoting member.

OTA has studies under way in nine program areas: energy and materials; industry, technology, and employment; international security and commerce; biological applications; food and renewable resources; health; communication and information technologies; oceans and environment; and science, transportation, and innovation.